The evolution and engineering of carbon fixation in plants



Jacques William Bouvier

St Cross College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2024

'Don't go mad, son.'
– Papa

Abstract

Photosynthesis converts atmospheric CO_2 into the sugars that sustain nearly all life on Earth. In plants, CO_2 diffuses into the leaf through stomata where it is subsequently assimilated by the enzyme rubisco. As the rate of CO_2 assimilation sets an upper limit on the rate of plant growth, CO_2 flux through both rubisco and stomata have long been identified as key bottlenecks that limit plant productivity. In this thesis, I aim to advance our understanding of the evolution of CO_2 assimilation in plants whilst also using genetic engineering to attempt to improve this process. This goal is achieved through two complementary approaches which focus on rubisco and stomata, respectively. In Chapter 2 I conduct a phylogenetically resolved investigation of rubisco catalytic evolution and find that trade-offs between enzyme kinetic traits are weaker than previously thought. I also make the novel discovery that phylogenetic constraints have played a more important role than kinetic trade-offs in limiting rubisco optimisation. In Chapter 3, I extend these phylogenetic analyses to investigate the rate of rubisco molecular and functional evolution. I find that rubisco is among the 1% of slowest-evolving genes on Earth. I also demonstrate that rubisco is not locked in evolutionary stasis but is continually evolving for improved catalysis and leaf-level photosynthesis in plants. The results from these two chapters have potential implications for rubisco engineering efforts and offer a new perspective to help understand the "rubisco paradox"–explaining why this enzyme appears moderately efficient at its primary function of fixing CO_2. In Chapter 4, I shift focus to stomata and show that stomatal density can be altered through targeted engineering of $NAD(P)^+$/NAD(P)H ratios in the model plant *Arabidopsis thaliana*. I also reveal that engineered plants exhibit improved photosynthetic water-use efficiency as well as enhanced vegetative growth under water-deficit conditions, without any impairment in net CO_2 assimilation rate or growth under optimal conditions. Together, this thesis thus harnesses insights from evolution and genetic engineering to improve our understanding of the past, present, and future potential of plant carbon fixation in nature as well as in applied engineering contexts.

Acknowledgements

As I type the final few words of this thesis, I am feeling sentimental. In my early days as a PhD student, I once wrote a piece for The Oxford Scientist publication in which I opened with the phrase '*Science is the pursuit of truth'*. Whilst somewhat pretentious in hindsight, what I meant by this at the time was that science provides us with the tools necessary to understand life in the world around us. After nearly half a decade since writing this phrase, and many ups and downs in my research along the way, I still firmly hold this sentiment close to my heart. However, what I have increasingly come to realise is the importance of the word *pursuit*. Science has led me on many remarkable journeys and has put me in touch with people I otherwise would not have met. I would like to emphasise here that it is these friends that I have gained, including a small number of individuals who are very special to me, which have made my time as a graduate student over these past years so memorable. Whilst I am certain that very few people will ever read these pages, I would like to acknowledge those who have been most important to either the production of this thesis or to my personal life as an individual. Upon reflection, these two groups were not always mutually exclusive. Though, sometimes it may have felt like they were.

First, to my supervisor Steve. You have been truly inspiring in many ways, and I am honoured to have been under your wing throughout this journey. You have a knack for making the complicated simple and you are ruthlessly efficient – an ideal recipe for any great scientist. If our work was ever under criticism (which it often was with countless rounds of peer-review), you could always cheer me up by retorting "*They just don't get it!*". Upon writing this acknowledgement, I am also humorously reminded of the time in which I was (very politely) banished from my lab bench until one of the manuscripts in this thesis materialised. I appreciate that you always kept me motivated and thinking about the big things, and that you regularly stopped me from chasing elusive rabbits down very deep holes. Most importantly, you always had faith in me, and that means a lot. We did have fun.

The next round of thanks go to the various colleagues and friends with whom I have shared the same dusty – and, later, mouse-infested – office, laboratory, or department space over the years. Whilst these many people are too numerous to mention here, it has been a pleasure to work and enjoy not

working with you all. Of this crowd, I am particularly indebted to Jane Langdale for her early and continued support, Lee Sweetlove for being an excellent second supervisor, and both Nick Harberd and Zhe Ji for hosting me as an initial rotation student. I have also been hugely fortunate to share my time as a graduate student with many wonderful members of the Kelly lab, including Rona, Oliver, Pippa, Lisa, Niels, Elizabeth, Jay, Florian, and Vijayalaxmi, as well as with the wider Langdale group who would often cheer us up over cake. My gratitude also extends to Spencer Whitney and his team at Australia National University, including Tim, Tanya, Davin, Hanjun, and Rosemary, for hosting me during a short sabbatical to undertake research that has not made it into this thesis. Whilst science can be a rather solitary pursuit, all of you people have made this not the case.

Now, the boys – Jochem, Jack, Jai, Jeremie, Nic, and Andrew – the girls – Lisa, Anya, and all partners past and present who appeared to tolerate us – and Thijs, for taking some steam out of Jochem on a regular basis. Over the years we have had many beers and even more laughs. I fondly cherish you all and the countless hours we have spent in each other's company. This shared company has been to the extent we have all become the same person. It's been a hell of a ride. You are my brothers for life. Raise your swords gentlemen.

Penultimately, to my parents, my grandparents, my brothers Lewis and Antoine, and the wider Bouvier and Hicks ensemble that have all been by my side from the start. You are the best family and support network that I could ask for. Your unconditional love, shelter, food, and monetary gifts over the years is appreciated beyond words. In fact, words could not do this justice. So, I will simply stress that I am grateful for everything. One day we will all share a cider at my apple orchard.

Last, but not least, Amy. Your kindness is unrelenting, your support is priceless, and your love knows no bounds. Perhaps a PhD in hindsight was not the wisest choice for someone with obsessive tendencies, and I acknowledge this must have not been fun for you at times. Similarly not fun for you was living hand to mouth for a very long time with scarce hot water or heating in our dark and damp flat up on Sidney Street. But, we always made it work. You are my best friend, and I could not have done it without you. Thank you.

Table of Contents

Abstract ... 4
Acknowledgements ... 6
Table of Contents ... 8
Chapter 1. General Introduction ... 16
1.1 Background ... 16
1.1.1 Food for thought – from field, to fork, to more forks, to fewer fields ... 16
1.1.2 Going through changes – future conditions will place increased burden on global crop production ... 17
1.1.3 On the road less travelled – improving photosynthesis can help feed the world ... 19
1.2 Photosynthesis ... 20
1.2.1 A primer ... 20
1.2.2 Photosynthesis on a primordial Earth ... 21
1.2.3 The emergence of oxygenic photosynthesis and the rise of photosynthetic eukaryotes ... 22
1.2.4 The 'light reactions' of photosynthesis ... 23
1.2.5 The 'dark reactions' of photosynthesis ... 24
1.2.6 C_3 photosynthesis in plants and the importance of rubisco and stomata ... 25
1.2.7 Other derived photosynthetic adaptations in plants ... 26
1.2.8 Opportunities and scope for optimising C_3 photosynthesis ... 26
1.2.9 Oppurtunities and scope for optimising rubisco and stomata ... 27
1.3 The form and function of rubisco in plant photosynthesis ... 28
1.3.1 Rubisco, the Earth's CO_2 fixing catalyst ... 28
1.3.2 The form and function of rubisco in photosynthetic and non-photosynthetic contexts ... 28
1.3.3 Rubisco is an imperfect catalyst ... 30
1.3.4 Rubisco suffers from additional limitations associated with its biogenesis, activation, and metabolic regulation ... 32
1.3.5 The rubisco paradox: a dichotomy between an important, yet inefficient, enzyme ... 34
1.3.6 The rubisco paradox has important implications for crop engineering ... 34
1.4 The form and function of stomata in plant photosynthesis ... 35
1.4.1 Physical constraints and the challenges of CO_2 diffusion in plants ... 35
1.4.2 Stomata control the exchange of CO_2 and water vapor in plants ... 36
1.4.3 Stomata dynamically modulate gas exchange in response to changing environments ... 36
1.4.4 Stomata modulate gas exchange across longer periods of time ... 38

1.4.5 Stomata present a key target for engineering enhanced plant performance 39
1.4.6 Stomatal development: knowledge gaps, outstanding questions, and future opportunities 41
1.5 Thesis overview 42
1.5.1 Thesis aims and objectives 42
1.5.2 Thesis layout 43
1.5.3 Collaboration 46
1.6 References 46
Chapter 2. Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off 70
2.1 Abstract 70
2.2 Introduction 71
2.3 Results 75
2.3.1 Rubisco kinetic data 75
2.3.2 A significant phylogenetic signal exists in rubisco carboxylase-related kinetic traits in angiosperms 75
2.3.3 Significant changes in rubisco kinetics occur during the evolution of C_4 photosynthesis 78
2.3.4 A significant phylogenetic signal exists in rubisco K_O in C_3 plants 80
2.3.5 Correlations between kinetic traits are weak in angiosperms and further relaxed after correcting for a phylogenetic signal 82
2.3.6 The evolution of rubisco kinetics is more limited by phylogenetic constraints than by catalytic trade-offs in angiosperms 85
2.3.7 Phylogenetic signal, weak kinetic trait correlations and strong phylogenetic constraint are features of rubisco in all photosynthetic organisms 87
2.4 Discussion 90
2. 5 Materials and Methods 95
2.5.1 Kinetic data 95
2.5.2 Phylogenetic tree inference 96
2.5.3 Phylogenetic signal analysis 97
2.5.4 Ancestral state estimation and mapping of kinetic traits to the phylogenetic tree 98
2.5.5 Least squares and linear regression models 98
2.6 References 99
Chapter 3. Rubisco is evolving for improved catalytic efficiency and CO_2 assimilation in plants 108
3.1 Abstract 108
3.2 Introduction 109
3.3 Results 114
3.3.1 RbcL has evolved slower than RbcS and has experienced stronger purifying selection . 114

3.3.2 RbcL is one of the slowest evolving genes in the tree of life 116
3.3.3 Rubisco is evolving for improved kinetic efficiency in plants 117
3.3.4 Rubisco is evolving for improved leaf-level CO_2 assimilation 120
3.4 Discussion 122
3.5 Materials and Methods 125
3.5.1 Rubisco sequence data 125
3.5.2 Rubisco phylogenetic tree inference 126
3.5.3 Quantification of the total extent of nucleotide and protein molecular evolution in rbcL/RbcL and rbcS/RbcS 127
3.5.4 Orthogroup classification and phylogenetic tree inference 127
3.5.5 Quantification of the percentile extent of rubisco molecular evolution within each taxonomic group 129
3.5.6 Integrated analysis of rubisco molecular and kinetic evolution 130
3.5.7 Integrated analysis of rubisco molecular evolution and CO_2 assimilation 132
3.6 References 134
Chapter 4. Metabolic engineering of stomatal precursor cells enhances photosynthetic water-use efficiency and vegetative growth under water-deficit conditions in Arabidopsis thaliana 144
4.1 Abstract 144
4.2 Introduction 145
4.3 Results 149
4.3.1 NOX can be expressed in plant cells and targeted to multiple subcellular locations 149
4.3.2 Identifying a promoter capable of driving stable and high expression of SmNOX in stomatal precursor cells 152
4.3.3 Validation of SmNOX expression and function 153
4.3.4 SmNOX activity causes reductions in stomatal density 154
4.3.5 SmNOX-induced changes in stomatal density improve plant photosynthetic water-use efficiency 156
4.3.6 SmNOX does not affect the kinetics of stomatal movement 158
4.3.7 SmNOX does not alter plant susceptibility to supra-optimal temperatures 160
4.3.8 SmNOX improves vegetative growth and biomass accumulation under water-deficit conditions 162
4.4 Discussion 164
4.5 Materials and Methods 168
4.5.1 Gene accession numbers 168
4.5.1 Gene synthesis 168
4.5.3 Construct design and assembly 169
4.5.4 Protoplast isolation, transformation, and visualisation 170

4.5.5 Plant material and stable transformation 170
4.5.6 Stomatal density measurements 170
4.5.7 Leaf mass area measurements 171
4.5.8 In vitro enzyme assays 171
4.5.9 Leaf gas exchange 172
4.5.10 Plant growth conditions 173
4.5.11 Plant growth assays 173
4.6 References 174
Chapter 5. General Discussion 188
5.1 Thesis key findings 188
5.2 Phylogeny provides a powerful framework to study enzyme function 189
5.3 The evolving landscape of rubisco: a revised biochemical–molecular constraint model helps to explain enzyme adaptation 196
5.3.1 Early evidence of constraints on rubisco adaptation 196
5.3.2 The seminal biochemical trade-off model posits a central compromise between specificity and CO_2 turnover 197
5.3.3 A revised trade-off model proposes all kinetic traits are antagonistically linked to form a one-dimensional landscape 198
5.3.4 Biochemical trade-offs revisited in light of increased data availability 199
5.3.5 Biochemical trade-offs revisited in light of phylogenetic signal 200
5.3.6 Rubisco phylogenetic signal – questions and advancements 202
5.3.7 Rubisco kinetic adaptations associated with C_3 versus C_4 photosynthesis 203
5.3.8 The relevance of C_3 versus C_4 kinetic adaptations and biochemical trade-offs 204
5.3.9 Why do rubisco kinetic trait differences exist between C_3 versus C_4 photosynthesis if not for biochemical trade-offs? 206
5.3.10 The discovery of a phylogenetic constraint 207
5.3.11 Phylogenetic constraints and the prevailing question of optimality 207
5.3.12 Phylogenetic constraints, molecular constraints, and the rubisco paradox 208
5.3.13 Future work needed to resolve molecular constraints 209
5.3.14 Potential implications for rubisco engineering 211
5.4 Metabolic engineering of stomatal density presents a potential future target for crop improvement 211
5.5 Concluding remarks 217
5.6 References 218
Appendix A, Supplemental Chapter: Response to Tcherkez and Farquhar: Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off 228
A.2 Introduction 229

A.3 Results ... 236
A.3.1 Kinetic and phylogenetic data ... 236
A.3.2 A phylogenetic signal cannot be generated with a randomly simulated arbitrary trait 236
A.3.3 A phylogenetic signal is a universal feature of arbitrary traits distributed according to Brownian motion ... 241
A.3.4 Phylogenetic signal is not an artefact of biases in species sampling ... 242
A.3.5 Phylogenetic signal in rubisco kinetics is not an artefact of using rbcL-based trees ... 243
A.3.6 Phylogenetic signal is not an artefact of method-to-method variability in trait measurements ... 244
A.3.7 Phylogenetic signal is not an artefact of C_3 vs C_4 photosynthesis ... 246
A.4 Discussion ... 248
A.5 Conclusions ... 251
A.6 Materials and Methods ... 252
A.6.1 Kinetic data ... 252
A.6.2 Phylogenetic tree inference based on rbcL and other gene sequences ... 253
A.6.3 Generation of randomly simulated arbitrary traits ... 254
A.6.4 Simulation of kinetic data on phylogenetic trees by Brownian motion ... 255
A.6.5 Phylogenetic signal analysis ... 255
A.7 Supplemental Tables ... 256
A.8 References ... 268
Appendix B, Supplemental Data and Text Supporting Chapter 2 ... 274
B.1 Supplemental Results ... 274
B.1.1 The phylogenetic signal in rubisco kinetic traits is not caused by overfitting due to use of rbcL for phylogenetic tree inference ... 274
B.2 Supplemental Figures ... 277
B.3 Supplemental Tables ... 289
B.4 References ... 295
Appendix C, Supplemental Data and Text Supporting Chapter 3 ... 296
C.1 Supplemental Results ... 296
C.1.1 Form I rubisco ancillary chaperones exhibit a variable rate of molecular evolution ... 296
C.1.2 Correlations in rubisco molecular and kinetic evolution are robust to biases in species sampling and model of sequence evolution ... 297
C.2 Supplemental Materials and Methods ... 298
C.2.1 Stratified sampling of rbcS sequences and phylogenetic tree inference ... 298
C.2.2 Genomes and gene models ... 298
C.2.3 Characterization of the set of enzymatic gene and protein sequences within orthogroups ... 299

C.2.4 Identification and classification of Calvin-Benson-Bassham cycle enzyme isoforms in land plants 299
C.2.5 Quantification of the relative extent of rubisco molecular evolution compared to all other Calvin-Benson-Bassham cycle enzymes in land plants 300
C.2.6 Quantification of the percentile extent of rubisco chaperone molecular evolution within each taxonomic group 301
C.2.7 Accounting for potential errors in the integrated analysis of rubisco molecular and kinetic evolution associated with species sampling and model of sequence evolution 302
C.3 Supplemental Figures 304
C.4 Supplemental Tables 309
C.5 References 325
Appendix D, Supplemental Data and Text Supporting Chapter 4 326
D.1 Supplemental Results 326
D.1.1 NOX expression results in tighter coordination between rates of assimilation and stomatal conductance 326
D.2 Supplemental Materials and Methods 327
D.2.1 Cloning and construct assembly 327
D.2.2 Vectors 327
D.2.3 Statistical analysis, model fitting, and parameter estimation of leaf gas exchange data 328
D.3 Supplemental Figures 330
D.4 Supplemental Tables 340
D.5 References 345

Further, however, perhaps.
For posterity…

Chapter 1. General Introduction

Chapter 1, Contributors:
Bouvier, J.W. and Kelly, S.

Chapter 1, Author contributions:
J.W.B. conceptualisation, reading, and writing. S.K. edits, comments, and discussion.

1.1 Background

1.1.1 Food for thought – from field, to fork, to more forks, to fewer fields

Agriculture is the cornerstone of modern civilisation. The cultivation of crops for food and animal feed paved the way for human societies to establish permanent settlements characterised by surplus food supplies, division of labour, and centralized governance (Murphy, 2007). Thousands of years later, little has changed. However, despite the enduring importance of agriculture, it would be incorrect to assume that the crops underpinning this practice emerged on Earth in an optimised state. Instead, these plants were domesticated and shaped from wild ancestors through a gradual process involving the selective propagation of desirable individuals, and their descendants, across a time-period spanning over 11,000 years (Miller, 2008). Thus, our agricultural crops are not evolutionarily static biological organisms. Instead, they have been consistently improved over the course of human history with far-reaching social, political, and economic implications for our species.

Although our elite crops are the product of intense breeding efforts sustained over millennia, the mid-20th Century marked unprecedented improvements in global food production (Khush, 1999). During this time-period, dubbed the Green Revolution (Evenson and Gollin, 2003), substantial advances in crop productivity were in large part driven by novel breeding initiatives (Smil, 2002; Hedden, 2003). Specifically, semi-dwarf and high-yielding varieties of wheat, rice, and maize were developed that exhibited enhanced resource allocation to grain, an increased responsiveness to nitrogen fertilization, and a greater resistance to lodging (Smil, 2002; Hedden, 2003). At the same time as

these breakthroughs in crop improvement, a transition toward intensive agriculture was facilitated by innovations in both the mechanisation of farm labour (Raza and Ameen, 2017) as well as the widespread availability of chemical fertilizers and pesticides (Smil, 2002; Kumar, 2007). Together, the combination of these technological advancements culminated in a two-fold increase in global cereal production over a 35-year-period and avoided millions of deaths from impending famines (Borlaug, 2003).

Despite the remarkable past success of the Green Revolution, trends in yield improvements have plateaued over recent decades (Hengsdijk and Langeveld, 2009; Ray *et al.*, 2012). As a consequence of this decline in the efficacy of international crop breeding initiatives, global food production is once again failing to keep pace with ever-rising demands from population growth and changing consumer diets (Ray *et al.*, 2013; Lowe, 2021; FAO *et al.*, 2024). Already, over three billion people suffer from malnutrition in the present-day, and an additional one billion people are deemed to be at critical risk of nutritional deficiencies based upon their financial circumstances (FAO, 2022). In future, this situation will become even more stark. By the mid-century, the 11 billion tons of food that we produce annually today will need to increase by at least two-fold or more in order to sustain the population (FAO, 2009, 2022; FAOSTAT, 2021). Coupled with the ongoing degradation of agricultural soils (Hossain *et al.*, 2020), and increasing competition for land for both urbanisation (Szabo, 2016; Kookana *et al.*, 2020) and biofuels (Tenenbaum, 2008), it is abundantly clear that we will need to produce food for more people on less space than has ever been achieved in history to realise food security over the coming century (Kelly, 2022).

1.1.2 Going through changes – future conditions will place increased burden on global crop production

Alongside the challenge of sustaining a growing population in the face of stagnating yields, climate change is poised to increasingly strain our future food production systems (Malhi, Kaur and Kaushik, 2021). For instance, since the dawn of the industrial era, atmospheric CO_2 has increased from 280 ppm in 1850 to over 420 ppm last year in 2023 (Betts *et al.*, 2024). Agriculture itself is in large part responsible for this change, contributing 8.5% of our total CO_2 annual emissions directly, as well as

an additional 14.5% of our annual emissions from secondary impacts associated with the destruction of natural habitats to create farm land (IPCC, 2019). Because greenhouse gases are a causal driver of global warming (Huang *et al.*, 2012; Kweku *et al.*, 2018; Al-Ghussain, 2019), our actions have already driven mean surface temperatures to increase beyond 1.5ºC above pre-industrial levels (IPCC, 2023; McCulloch *et al.*, 2024) thereby surpassing the threshold set in 2015 by The Paris Agreement international treaty on climate change (The United Nations, 2015). It is widely understood that in a world which exceeds this 1.5°C temperature warming scenario, the impacts and severity of climate change will be substantially increased (The United Nations, 2015) and an irreversible and catastrophic point-of-no-return may be reached with the onset of self-amplifying positive climate feedback loops (Ripple *et al.*, 2023). Alarmingly, our annual emission are also showing no signs of slowing and are still continuing to rise each year (Le Quéré *et al.*, 2016). On our current trajectory, atmospheric CO_2 levels are projected to exceed 550 ppm by the turn of the century (Stocker *et al.*, 2013) and we are on course to surpass 2°C warming above pre-industrial levels in the next decade (IPCC, 2023; McCulloch *et al.*, 2024). Thus, we are at the foot of an impending climate emergency.

Agricultural-based risk assessments predict that global warming will cause an increase in the severity, frequency, and duration of extreme weather events (Jones and Mearns, 2003). Moreover, there is also increased likelihood that simultaneous extreme weather events will occur and result in complete crop failures on a global scale (Kornhuber *et al.*, 2023). Of all projected changes in weather patterns, the combination of rising surface temperatures, increased frequency of heatwaves, and altered precipitation patterns are amongst the most concerning for agriculture as would drive prolonged and more intense periods of drought (Dai, Zhao and Chen, 2018). This is particularly alarming because water scarcity is already one of the biggest threats to our food production systems contributing toward harvest losses in over 40% of global farmlands (Liu *et al.*, 2022). As the effects of anthropogenic climate change become exacerbated, it is predicted that this figure will to rise to beyond 60% by 2050 (Liu *et al.*, 2022). Whilst irrigation can help partially mitigate these impacts of water deficit on agriculture, this requires costly infrastructure at a scale of investment and human resources that is not feasible across much of the developing world (Kirpich, Haman and Styles, 1999; Plusquellec, 2002). Moreover, reliance on irrigation is also not a long-term solution as it places

increased burden on limited groundwater reserves (Siebert *et al.*, 2010) and if poorly managed, can lead to further complications associated with salinization of agricultural soils (Saysel and Barlas, 2001). Therefore, not only will we be required to produce more food on less land than ever before, but we will also have to do this in the face of a more unpredictable and harsh set of circumstances. Climate change presents a major humanitarian challenge with drastic implications for our global food security.

1.1.3 On the road less travelled – improving photosynthesis can help feed the world

Given that agriculture will not meet forecast production demands, and as our food production systems are set to become increasingly threatened by climate change (Ray *et al.*, 2013), new technologies are urgently needed to help feed the world in the 21st century. In particular, novel innovations are required to unlock the latent genetic potential of our crops in order to push productivity gains beyond the current capabilities of conventional breeding. In parallel, plant science research will also be tasked with developing crops that are better equipped to tolerate adverse climate conditions without compromising upon performance or yield. Crucially, the solutions underpinning both of these goals must prioritize environmental sustainability to ensure the long-term health of our agricultural systems.

Photosynthesis constitutes one promising target for crop improvement because this process provides all of the energy which fuels plant growth and development. Importantly, photosynthesis also presents a largely untapped reservoir for potential optimisation because it is surprisingly inefficient in most plants (including crops) (Zhu, Long and Ort, 2010; Long, Marshall-Colon and Zhu, 2015). For instance, only 4.6% of total incident solar energy is converted into organic biomass in the vast majority of plants which utilize the C_3 mode of photosynthesis (Zhu, Long and Ort, 2008). Moreover, measured rates of photosynthetic conversion efficiency can be further reduced to ~2% in some crops under real-world field conditions (Monteith, 1977). In large part, this lack of optimisation in crop photosynthesis is the consequence of breeding strategies which have historically selected for enhanced yield by manipulating carbon partitioning, rather than by directly increasing the assimilation of this organic carbon through photosynthesis (Richards, 2000). In some cases, this historic legacy has even led to reported examples of weak or absent correlations between

photosynthetic rates and the harvest index of crops (Evans and Dunstone, 1970; Evans, 1997; Long *et al.*, 2006) despite considerable evidence that crop productivity and yields are directly related to photosynthetic efficiency (Fischer *et al.*, 1998; Lilley, Bolger and Gifford, 2001; Miglietta *et al.*, 2001; Calfapietra *et al.*, 2003; Ainsworth and Long, 2005, 2021; Gielen *et al.*, 2005; Kruger and Volin, 2006; Long *et al.*, 2006; Zhu, Long and Ort, 2010; Weigel and Manderscheid, 2012; Simkin, López-Calcagno and Raines, 2019). Together, improving the efficiency of photosynthesis is thus considered to present a promising and overlooked avenue for crop improvement.

In light of the potential of photosynthesis for increasing crop yields, identifying strategies to optimise this process represents an important goal of future research. As photosynthesis is central to all aspects of physiology at the intersection of plants and their environment, genetic manipulation of this process also offers considerable scope to develop more climate-resilient crops. Of key significance, engineering photosynthesis will offer handles to optimise plant-water relations and increase crop tolerance to future water shortages and drought conditions. In summary, enhancing photosynthesis using precision breeding or biotechnological engineering may thus provide important solutions to help feed the world in the 21st century without destroying the planet. As food security represents one of the most pressing challenges which threatens our species, research advancing our understanding and engineering of photosynthesis could therefore possibly help contribute toward a more sustainable future for humanity.

1.2 Photosynthesis

1.2.1 A primer

Photosynthesis is the process which assimilates atmospheric CO_2 into the sugars that sustain nearly all life. More precisely, photosynthesis is a collective term that encompasses all of the disparate light-dependent and light-independent biochemical reactions which function to convert inorganic carbon into organic biomass. By orchestrating this remarkable feat, photosynthesis represents the cornerstone of biology and operates at the intersection between the atmosphere and biosphere of our planet.

1.2.2 Photosynthesis on a primordial Earth

Given its fundamental role at the basis of the global food chain, understanding the origin of photosynthesis holds important implications for our appreciation of life on Earth. Accordingly, this topic represents a key research area in biology which has long captivated scientists (Hohmann-Marriott and Blankenship, 2011). However, despite having emerged only once in a single-celled bacterium, re-tracing the early evolution of photosynthesis is convoluted by both the incomplete biogeochemical record (Schopf, 2011), as well as historic lateral gene transfer events which have meant that all extant major phyla containing photosynthetic bacteria today also contain non-phototrophic sister species in the same clade (Fischer, Hemp and Johnson, 2016). In this way, although there is an innate human impulse to seek beginnings, the beginning of photosynthesis has remained enigmatic with several competing hypotheses having been proposed (Mulkidjanian *et al.*, 2006; Sadekar, Raymond and Blankenship, 2006; Tice and Lowe, 2006; Cardona, 2019).

To the best of our current understanding, anoxygenic photosynthesis was the first form of photosynthesis that emerged ~3.5 billion years ago during the early history of our planet (Schidlowski, 1988; Tice and Lowe, 2004, 2006; Hohmann-Marriott and Blankenship, 2011; Czaja *et al.*, 2013). Since its discovery (Molisch, 1907), many organisms capable of anoxygenic photosynthesis are now characterised that are distributed broadly across the bacterial and archaeal domains of life (Yabe *et al.*, 2022). Curiously, these extant taxa are often found in extreme conditions, and thrive in some of the harshest locations of our planet including volcanic springs, frozen sea ice, hypersaline or soda lakes, and hydrothermal vents (Imhoff, 2017). Thus, anoxygenic photosynthesis is an ancient and robust mode of autotrophy. The organisms which used this metabolism emerged under the hostile conditions of our primordial Earth, and they continue to push the boundaries of life in all directions today.

Consistent with their remarkable ecological diversity, anoxygenic phototrophs also exhibit considerable variability in both their physiology and metabolism. Importantly, the only single definition that appears to unify this mode of photosynthesis is that it does not utilize water as an electron donor and that it does not evolve molecular oxygen as a waste product (Hanada, 2016; George, Vincent and Mackey, 2020). Instead, anoxygenic photosynthesis uses diverse geochemically derived

reductants or organic compounds to provide electrons (Hanada, 2016; George, Vincent and Mackey, 2020). Moreover, molecular oxygen is in fact harmful to many anoxygenic phototrophs (Buchanan, 1992). Despite the many unanswered questions surrounding anoxygenic photosynthesis, it is evident that this pathway provides a window into the origins of photosynthesis as well as into the early life on our planet.

1.2.3 The emergence of oxygenic photosynthesis and the rise of photosynthetic eukaryotes

Following the emergence of anoxygenic photosynthesis, oxygenic photosynthesis is thought to have evolved as a derived trait in an ancient cyanobacterium (Sekine *et al.*, 2011; Williford *et al.*, 2011; Johnson *et al.*, 2013; Fischer, Hemp and Johnson, 2016) that lived at least 2.4 billion years ago (Hofmann, 1976; Lyons, Reinhard and Planavsky, 2014; Planavsky *et al.*, 2014; Cardona, Murray and Rutherford, 2015; Fischer, Hemp and Johnson, 2016). In contrast to its ancestral form, the underlying molecular machinery and apparatus which underpins oxygenic photosynthesis is considerably more complex and required several key innovations to occur. Specifically, this includes the novel acquisition of two photosystems I (PSI) and II (PSII) that are linked in parallel, as well as the remarkable gain-of-function evolution of PSII to contain the only water-oxidizing complex (water-plastoquinone oxidoreductase) in biology capable of catalysing the four-electron oxidation of water to molecular oxygen (Duysens, Amesz and Kamp, 1961; Witt, Mueller and Rumberg, 1961; Cardona, Murray and Rutherford, 2015; Fischer, Hemp and Johnson, 2016). Together, these two adaptations represent the hallmarks of oxygenic photosynthesis and enable modern autotrophs to generate biomass and oxygen across the entire surface of our globe from only CO_2, water, and light.

In accordance with its ecological significance in our biosphere, the evolution of oxygenic photosynthesis represented a landmark event that had pervasive geochemical impacts on our planet (Shaw, 2016). As water is highly abundant on Earth, the novel capacity to utilize this respective molecule as an electron source fuelled an unprecedented spike in global primary production (Ward, Rasmussen and Fischer, 2019). Importantly, this led to The First Great Oxidation Event (Eigenbrode and Freeman, 2006; Koch and Britton, 2008) during which atmospheric oxygen concentrations

increased by four orders of magnitude in a blink of geological time (Cloud, 1973; Walker *et al.*, 1983; Farquhar James, Bao Huiming and Thiemens Mark, 2000; Bekker *et al.*, 2004; Papineau, Mojzsis and Schmitt, 2007; Kump, 2008; Guo *et al.*, 2009; Hoffman, 2013; Johnson *et al.*, 2013, 2014; Rasmussen, Bekker and Fletcher, 2013). Concomitant with this change in O_2, the rapid increase in photosynthetic activity also underpinned a decline in atmospheric CO_2 concentrations driven by both the sequestration and long-term burial of organic biomass (Zhao *et al.*, 2023) and associated enhanced rates of chemical weathering (Kasting, Toon and Pollack, 1987). Thus, oxygenic photosynthesis transformed the abiotic conditions of Earth from a highly reducing environment into the aerobic world of the present day.

As a consequence of altering the redox state of our planet, the emergence of oxygenic photosynthesis also fundamentally changed the trajectory of life. Specifically, this pivotal event fuelled the evolution of aerobic respiration, eukaryotes, and the rise of multicellular life in the “big bang of evolution” (Barber, 2004). Subsequently, this biological complexity on Earth was further expanded ~1.5 billion years ago when the ability to perform photosynthesis was transferred to eukaryotes through the endosymbiosis of a cyanobacterium (Schwartz and Dayhoff, 1978; Hedges *et al.*, 2004). This event gave rise to autotrophic eukaryotes containing endosymbiotic chloroplasts, and created the lineage extending toward the plants and algae that dominate primary production along with cyanobacteria in our current biosphere (Delwiche, 1999; Keeling, 2010; Archibald, 2015). Together, it thus follows that oxygenic photosynthesis is the engine of biology. Its evolution not only transformed the conditions of our atmosphere and oceans, but it also fuelled the evolution of the life which flourishes within it.

1.2.4 The ‘light reactions’ of photosynthesis

At the biochemical level, oxygenic photosynthesis is comprised of a complex series of reactions which are co-ordinated between two distinct but parallel phases. The first of these phases comprise the photochemical reactions which serve to convert solar radiation into chemical energy. Owing to the dependence of these reactions on light, these are often termed the light-dependent reactions (or ‘light reactions’) of photosynthesis. For an excellent recent review of the light reactions, see (Johnson, 2016).

In brief, in the dominant non-cyclic pathway, the water-oxidizing complex of PSII first splits H_2O into its constituent parts of oxygen, protons, and electrons in the following stoichiometry $2H_2O \rightarrow O_2 + 4H^+ + 4e^-$. Next, these electrons are accepted by PSII and are excited to a higher energy state by light absorbed at a central chlorophyll within the light harvesting complex. These excited electrons are then funnelled from PSII along a redox gradient spanning plastoquinone cofactors and the cytochrome b6f complex to PSI. Finally, these electrons are excited once more before being transferred via a ferredoxin to the terminal electron acceptor nicotinamide adenine dinucleotide phosphate $(NADP)^+$ to generate NADPH. In parallel to the above electron transfer reactions, an electrochemical gradient is produced between the thylakoid lumen and either the chloroplast stroma (eukaryotes) or the cell cytoplasm (cyanobacteria) (Johnson, 2016). This is because coupled with the release of electrons, the water-splitting activity of PSII in the thylakoid lumen also produces oxygen that is liberated as a by-product, as well as protons which cannot escape and thus accumulate to high concentrations. Consequently, this creates a charge separation which is harnessed to generate adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and Pi by the downhill movement of protons along their electrochemical gradient through a thylakoid membrane-bound ATP synthase (Nelson and Ben-Shem, 2004; Nelson and Yocum, 2006). Due to the combined action of these electron and proton transfer steps, the light reactions of photosynthesis thus ultimately harness solar energy to drive the production of ATP and NADPH from the oxidation of water.

1.2.5 The 'dark reactions' of photosynthesis

After generation of ATP and NADPH, these respective energy storage molecules are subsequently used to power the downstream light-independent carbon fixation reactions (or 'dark reactions') of photosynthesis. Specifically, this secondary biochemical phase of photosynthetic metabolism is organised into an autocatalytic series of enzyme-mediated steps that together comprise the Calvin-Benson-Bassham reductive pentose phosphate pathway (Bassham *et al.*, 1954). Crucially, the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) catalyses the first reaction in this scheme, where it fixes atmospheric CO_2 onto the acceptor molecule ribulose-1,5-bisphosphate (RuBP) to produce two molecules of 3-phosphoglycerate (3-PGA). After this rubisco-mediated

carboxylation reaction, 3-PGA molecules then become chemically reduced by downstream metabolic enzymes in the pathway to produce carbon skeletons. At this point, a key fork in the Calvin-Benson-Bassham cycle is reached. These carbon skeletons can either remain in the cycle to re-generate the initial substrate RuBP and allow rubisco-mediated carbon fixation to continue, or alternately they can be exported to produce the organic sugars which underpin aerobic metabolism. In this way, the light and dark reactions of photosynthesis work in concert to convert solar energy into the organic sugars which drive nearly all cellular metabolic processes.

1.2.6 C_3 photosynthesis in plants and the importance of rubisco and stomata

Whilst there are a number of photosynthetic pathways that have emerged across the tree of life, the vast majority of all plants utilize a mode of photosynthesis termed C_3 photosynthesis. Specifically, this pathway is so-called because the initial organic compound into which CO_2 is assimilated is the three-carbon metabolite 3-PGA. In C_3 plants, this situation arises because atmospheric CO_2 is delivered to the site of rubisco-mediated carbon fixation by simple diffusion. More precisely, CO_2 enters the leaf through stomatal pores on the leaf epidermis and moves along its concentration gradient into the chloroplast stoma of the mesophyll where it is assimilated by rubisco.

Enzyme collision theory dictates that the rate of a given reaction is dependent upon the likelihood of collisions between a catalyst and its substrate (Bartholomay, 1962; Keleti and Welch, 1984). According to this principle, it follows that both the stomata and rubisco represent key control points which determine the efficiency of photosynthetic carbon fixation in C_3 plants. This is because stomata mediate the diffusion of CO_2 to the site where rubisco is localised and thus control the substrate concentration around this enzyme. For a given substrate concentration, rubisco then catalyses the assimilation of CO_2 at a rate which is dependent upon its inherent catalytic competency.

Despite their importance, both rubisco and stomata exhibit inherent limitations in their function that constrain C_3 photosynthesis. At the site of stomata, a trade-off arises because water vapour is simultaneously lost from the plant through transpiration along the inverse route to CO_2 entry due to the opposing concentration gradients of both respective CO_2 and H_2O gases. Thus, this means that plants must balance a fundamental compromise between carbon gain and water loss in response to

changes in physiological demands as well as external conditions. In addition to stomata, rubisco is a relatively inefficient enzyme which exhibits a modest rate of CO_2 turnover (Bar-Even *et al.*, 2011) and also catalyses a secondary reaction with O_2 which competitively inhibits the enzyme's active site and also drives the loss of previously fixed CO_2 through photorespiration (Bowes, Ogren and Hageman, 1971; Lorimer, Andrews and Tolbert, 1973). Together, these shortcomings of rubisco catalysis limit its CO_2 assimilation capacity and ultimately the net flux at which carbon can become fixed by photosynthesis. Thus, both stomata and rubisco present significant bottlenecks in carbon fixation in C_3 plants which curtail the productivity of this process over time and space.

1.2.7 Other derived photosynthetic adaptations in plants

In addition to C_3 photosynthesis, derived photosynthetic adaptations have also evolved in nature which deliver CO_2 to rubisco through mechanisms other than diffusion. Collectively, these pathways are termed CO_2 concentrating mechanisms, and in plants include C_4 photosynthesis and crassulacean acid metabolism (CAM). In both of these respective pathways, supra-atmospheric CO_2 concentrations are maintained in the mesophyll chloroplast through incorporation of an additional CO_2 fixation step that is either spatially (C_4) (Sage, Sage and Kocacinar, 2012) or temporally (CAM) (Dodd *et al.*, 2002) separated from the primary rubisco-mediated CO_2 fixation reaction. To achieve this chemical feat, the oxygen-insensitive enzyme phospho*enol*pyruvate carboxylase first assimilates CO_2 into a four-carbon compound which is then transported and subsequently decarboxylated in the vicinity of rubisco to act like a biochemical CO_2 pumping mechanism. In this way, both of these derived biochemical pathways enhance the efficiency of photosynthetic CO_2 fixation. Importantly, this improvement is also achieved without any requirement to either alter the inherent catalytic competency of rubisco or to increase transpiration through the stomata.

1.2.8 Opportunities and scope for optimising C_3 photosynthesis

Although C_4 photosynthesis and CAM effectively overcome important constraints on C_3 photosynthesis as is demonstrated by their considerable ecological success (Still *et al.*, 2003; Lüttge, 2004), both pathways are exceedingly difficult to reproduce with synthetic biology. This is because the co-ordination of these photosynthetic metabolisms depend upon a complex series of adjustments

to plant anatomy, morphology, biochemistry, and ultrastructure (Edwards, 2019; Schluter and Weber, 2020). Accordingly, this has meant that efforts which have sought to engineer these complex traits in order to enhance the photosynthetic productivity of C_3 plants have, to date, been met with limited success (DePaoli *et al.*, 2014; Ermakova *et al.*, 2020; Furbank, Kelly and von Caemmerer, 2023; Yang *et al.*, 2024).

As derived photosynthetic pathways do not present tractable targets for crop improvement, future engineering efforts should focus on discovering more accessible handles for achieving photosynthetic improvements (*'Don't reinvent the wheel, just realign it*'.) These efforts might prioritise the key control-points that represent important determinants of carbon fixation efficiency in C_3 photosynthesis. First, the stomata which modulates photosynthetic capacity by determining the rate of CO_2 supply to rubisco. Second, rubisco which modulates photosynthetic capacity by determining the rate at which CO_2 can be converted into organic sugars. Accordingly, it follows that engineering improvements in either rubisco or stomata present a promising avenue to enhance crop productivity. In support of this application, it also follows that research seeking to elucidate the evolution, adaptation, and functional limitations of both rubisco and stomata will provide important perspectives to help inform future engineering strategies.

1.2.9 Opportunities and scope for optimising rubisco and stomata

To lay the foundation for the research contained with this thesis, our current knowledge of the form and function of rubisco and stomata are discussed in depth in the dedicated introduction sections below. The ultimate purpose of this extended introduction is to highlight outstanding questions associated with the constraint of rubisco and stomata in extant C_3 plants, with a view to outline the challenges and opportunities for their potential improvement in future. Following the order in which they evolved on Earth, rubisco is discussed first in Chapter 1.3, followed by stomata in Chapter 1.4.

1.3 The form and function of rubisco in plant photosynthesis

1.3.1 Rubisco, the Earth's CO_2 fixing catalyst

Ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco, EC 4.1.1.39) is the enzyme that assimilates atmospheric CO_2 into the organic molecules that underpin life on Earth. Whilst several metabolic pathways can facilitate this rubisco-mediated reaction, the vast majority of carbon enters our biosphere through the enzyme's role in the Calvin-Benson-Bassham cycle of oxygenic photosynthesis. In this scheme, rubisco catalyses the assimilation of atmospheric CO_2 onto ribulose-1,5-bisphosphate (RuBP) to produce the building blocks and precursors that fuel nearly all biological processes. In this way, rubisco sustains life on our planet and has produced nearly all organic biomass that has ever existed.

1.3.2 The form and function of rubisco in photosynthetic and non-photosynthetic contexts

Since the initial divergence of *bona fide* rubisco from non-CO_2 fixing rubisco-like proteins (Erb and Zarzycki, 2018), diverse lineages of this enzyme have evolved across all three domains of life including the archaea, prokaryotes and eukaryotes (Anderson, Price and Fuller, 1968; Ezaki *et al.*, 1999; Watson, Yu and Tabita, 1999). Among these, three major rubisco variants exist called Form I, Form II, and Form III. Whilst Form III rubisco participates in ribonucleoside metabolism in archaea and some bacterial phyla (Sato, Atomi and Imanaka, 2007; Aono *et al.*, 2015; Wrighton *et al.*, 2016; Jaffe *et al.*, 2019), both Form I and II rubisco have been recruited for oxygenic photosynthesis. Though, Form I rubisco underpins over 90% of global CO_2 assimilation and is responsible for sequestering ~100 gigatons of carbon annually (Field *et al.*, 1998). Accordingly, Form I rubisco is also the most abundant rubisco variant and is also the most abundant enzyme on Earth with a combined total mass of ~0.7 gigatonnes (Ellis, 1979; Bar-On and Milo, 2019). Thus, Form I rubisco is the predominant CO_2 fixing enzyme on our planet which catalyses the exchange of carbon at the intersection of Earth's atmosphere and biosphere.

In nature, Form I rubisco is found across autotrophic taxa with diverse physiologies and ecologies, and can be subdivided into four main phylogenetic lineages coined Form IA – ID (Tabita *et al.*, 2008).

Form IA and Form IB represent the 'green-type' rubisco lineages which are found in both proteobacteria and α-cyanobacteria (Form IA), as well as plants, green algae, and β-cyanobacteria (Form IB). In contrast, Form IC and Form ID represent the 'red-type' rubisco lineages that occur in proteobacteria (Form IC) as well as in non-green algae (Form ID). This broad distribution across the tree of life is consistent with the ancient origin of the Form I lineage which dates back ~2.4 billion years (Kumar *et al.*, 2017).

Despite the considerable phylogenetic diversity and complex evolutionary history of Form I rubisco, this enzyme exhibits a surprising degree of conservation at the level of both its protein sequence and structure. Notably, all Form I rubiscos are comprised of eight ~50 kDa rubisco large (RbcL) subunit proteins and eight 12–18 kDa rubisco small (RbcS) subunit proteins to form an L_8S_8 hexadecamer (Schneider, Lindqvist and Brändén, 1992; Tabita *et al.*, 2008). This complex oligomeric arrangement distinguishes Form I from all other rubiscos (including rubisco-like protein homologs) that are exclusively composed of either single or repeating RbcL homodimers. Instead, the structure of Form I rubisco contains four RbcL dimers $(L_2)_4$ organised into an octameric ring that is capped by a tetrad of RbcS proteins at both the top and bottom. As the single conserved unit among all of these rubiscos, the RbcL dimer is arranged head-to-tail (i.e., antiparallel) such that two active sites are formed at the interface of the N-terminal α/β domain of one RbcL protein and the C-terminal $(\beta/\alpha)_8$ triosephosphate isomerase (TIM) barrel of the adjacent RbcL subunit. In this way, RbcL encodes the active site residues and represents the minimal ancestral subunit required for catalysis (Andrews, 1988; Lee and Tabita, 1990).

In contrast to RbcL, RbcS was recruited at the dawn of the Form I lineage and is essential for the structure and solubility of this variant (Andrews, 1988; McFadden and Small, 1988; Fitchen *et al.*, 1990; Lee, Berka and Tabita, 1991; Genkov and Spreitzer, 2009; Liu *et al.*, 2010; Grabsztunowicz *et al.*, 2015; Joshi *et al.*, 2015). Despite its distance from the active site, incorporation of RbcS is also believed to have underpinned the superior catalytic efficiency of Form I rubisco by increasing the biochemical landscape accessible for its adaptation (Schulz *et al.*, 2022). Whilst there are many scenarios in which useless or non-adaptive complexity may arise during evolution (i.e., the accumulation of mutations which are neutral in a complex state but become deleterious in a simple

state) (Schulz, Sendker and Hochberg, 2022), recruitment of RbcS thus appears to represent a *bona fide* adaptation which improves fitness. This important role is further demonstrated when RbcS is either mutated (Voordouw *et al.*, 1987; Read and Tabita, 1992b; Kostov, Small and McFadden, 1997; Spreitzer and Salvucci, 2002), or where chimeric holoenzymes have been created by substituting whole RbcS between species (Andrews, Greenwood and Yellowlees, 1984; Andrews and Lorimer, 1985; van der Vies, Bradley and Gatenby, 1986; Lee, Read and Tabita, 1991; Read and Tabita, 1992a; Getzoff *et al.*, 1998; Horken and Tabita, 1999; Kanevski *et al.*, 1999; Wang *et al.*, 2001; Karkehabadi *et al.*, 2005; Ishikawa *et al.*, 2011; Matsumura *et al.*, 2020; Sakoda *et al.*, 2021). It is also possible that RbcS may have a latent physiological role in regulating enzyme function because both environmentally-driven (Cavanagh, Slattery and Kubien, 2022) and organ-specific (Morita *et al.*, 2014; Lin *et al.*, 2020; Martin-Avila *et al.*, 2020) variability in plant isoform expression has been found to impact holoenzyme kinetics. Thus, RbcL is the catalytic subunit and the primary determinant of kinetic variation across Form I variants. Though, both RbcL and its cognate RbcS are important components of enzyme function and fitness.

1.3.3 Rubisco is an imperfect catalyst

In spite of its global abundance and importance, Form I rubisco is far from perfect. In fact, the prevailing view of the enzyme is that it is a poorly efficient and mal-optimised catalyst. Whilst there has been some recent defence of rubisco as being 'not so bad' (Bathellier *et al.*, 2018), its more widely-held reputation of being an inefficient enzyme stems from two key catalytic shortcomings as well as a number of other peculiarities. Together, these collective limitations constrain rubisco's CO_2 assimilation capacity and thereby limit flux through photosynthesis.

First, all Form I rubisco catalyse a rate of CO_2 turnover of $<12\ s^{-1}$ which is sub-optimal in the context of the average enzyme (Bar-Even *et al.*, 2011) and lags by several orders of magnitude behind the fastest enzymes on Earth which surpass $10^5\ s^{-1}$ at the diffusion limit of catalysis (Albery and Knowles, 1976; Bar-Even *et al.*, 2011). Whilst this sluggish rubisco reaction can in part be attributed to the energetic and thermodynamic barriers of the carboxylation reaction (Bar-Even *et al.*, 2012; Cotton, Edlich-Muth and Bar-Even, 2018; Prywes *et al.*, 2022), rubisco is not even among the most efficient CO_2-fixing enzymes on our planet. Instead, many other carboxylases (including

phospho*enol*pyruvate (PEP) carboxylase, PEP carboxykinase, pyruvate carboxylase, and crotonyl-CoA carboxylase/reductase) as well as several CO_2 reducing enzymes (including formate dehydrogenase) catalyse considerably faster turnover rates than rubisco (Bar-Even *et al.*, 2011; Schuchmann and Müller, 2013; Maia *et al.*, 2016; Cotton, Edlich-Muth and Bar-Even, 2018). Moreover, this moderate turnover is further exacerbated in the real-world in which the time-averaged rubisco rate of reaction is 100-fold slower in terrestrial habitats compared to those measured *in vitro* (Bar-On and Milo, 2019). Thus, rubisco is a slow enzyme that is far from the physiochemical limits of catalysis, and is very slow when considering its importance to host fitness in the context of other enzymes.

In addition to being slow, all known rubiscos discovered to date also catalyse a competitive reaction with O_2 in addition to their primary CO_2 reaction (Bowes, Ogren and Hageman, 1971; Lorimer, Andrews and Tolbert, 1973). This inability to discriminate CO_2 and O_2 is thought to result from the inherent physical and chemical limitations of CO_2 as a substrate (including its inertness, hydrophobicity, and low molecular mass), as well as the similarities in the size and electrostatic potential of both featureless CO_2 and O_2 gaseous molecules (Bathellier *et al.*, 2018). Although this reaction promiscuity is not necessarily an uncommon property among enzymes in nature (O'Brien and Herschlag, 1999; Copley, 2003), in the case of rubisco it results in a substantial burden to host productivity. This is because whereas rubisco's carboxylation reaction produces two molecules of 3-PGA which enter the photosynthetic Calvin-Benson-Bassham cycle, rubisco's competitive oxygenation reaction produces only one molecule of 3-PGA in addition to a molecule of 2-phosphoglycolate (2-PG). Moreover, this 2-PG product is toxic (inhibiting key enzymes in photosynthesis, glycolysis, and starch metabolism (Anderson, 1971; Kelly and Latzko, 1976; Hall *et al.*, 1987)), wasteful (diverting carbon and nitrogen away from the Calvin-Benson-Bassham cycle (Wingler *et al.*, 2000)), and produced in excessively high quantities (with oxygenation representing around 1/3 of all rubisco reactions under present-day conditions (Sharkey, 1988; Walker *et al.*, 2016)) such that its downstream recycling is essential. However, energetic costs arise because metabolising a molecule of 2-PG via the photorespiratory salvage pathway releases 0.5 molecules each of previously fixed CO_2 and NH_3, and also consumes 2 NADH as well as 3.5 ATP (Walker *et*

al., 2016) (with this cost rising to 12.25 ATP upon re-fixation of the released CO_2 and NH_3) (Peterhansel *et al.*, 2010). Thus, although photorespiration serves several beneficial functions in autotrophs (Bloom and Lancaster, 2018; Busch, Sage and Farquhar, 2018; Eisenhut, Roell and Weber, 2019; Busch, 2020; Jiang *et al.*, 2023), rubisco's oxygenase activity considerably limits photosynthetic productivity by competitive inhibition of the active site as well as by draining previously fixed carbon and energy.

1.3.4 Rubisco suffers from additional limitations associated with its biogenesis, activation, and metabolic regulation

Alongside being a slow and error-prone catalyst, rubisco also suffers from numerous additional shortcomings beyond its kinetics that serve to constrain enzyme function. In particular, these alternate limitations are instead associated with the enzyme's complex and costly regulatory mechanisms which modulate its biogenesis, activation, and metabolic function

First, Form I holoenzyme biogenesis does not occur spontaneously. Instead, the enzyme exhibits complex folding and assembly requirements which depend upon an extensive network of auxiliary chaperones. Specifically, this includes the chloroplast Chaperonin (Cpn) Cpn60α/β and Cpn20/10, Rubisco accumulation factor 1 and 2 (Raf1 and Raf2), Bundle sheath defective 2 (BSD2), and Ribulose-bisphosphate-carboxylase factor X (RbcX). Our understanding of the exact contribution of each of these components and the steps that they mediate has been the subject of a number of recent reviews (Bracher *et al.*, 2017; Hayer-Hartl, 2017; Wilson and Hayer-Hartl, 2018; Hayer-Hartl and Hartl, 2020; Gionfriddo, Zang and Hayer-Hartl, 2023), and represents the culmination of considerable research effort that has been sustained over several decades (Brutnell *et al.*, 1999; Saschenbrecker *et al.*, 2007; Bracher *et al.*, 2011; Tsai *et al.*, 2012; Feiz *et al.*, 2012, 2014; Wheatley *et al.*, 2014; Doron *et al.*, 2014; Kolesinski *et al.*, 2014; Hauser *et al.*, 2015; Aigner *et al.*, 2017; Wilson *et al.*, 2018; Yan *et al.*, 2018; Zhao and Liu, 2018; Conlan *et al.*, 2019). In plants, this process first comprises chloroplast Cpn60α/β and their Cpn20/10 co-factors (homologs of bacterial GroEL and GroES, respectively) to encapsulate individual RbcL polypeptide chains and support their correct folding in the absence of aggregation. Either in parallel or in tandem to one another, Raf1 and RbcX then dynamically bind these folded RbcL proteins to support formation of L_2 dimers, and

subsequently, formation of the $(L_2)_4$ octameric core. Finally, BSD2 binds to RuBP catalytic sites to maintain the L_8 structure in an inactive state which is amenable to recruitment of RbcS, whose folding is mediated after chloroplast import by Hsp70. In addition to the above steps, Raf2 is also known to have an essential, but as yet undetermined role in the formation of the $(L_2)_4$ core (Aigner *et al.*, 2017) and possibly also in the interaction and recruitment of RbcS (Feiz *et al.*, 2014; Wheatley *et al.*, 2014). Thus, Form I rubisco biogenesis is complex.

Second, Form I rubisco is not directly functional after assembly. Instead, its catalytic competency depends upon an initial activation step in which a non-substrate CO_2 molecule attaches to (i.e., 'carbamylates') a lysine-201 residue causing the Mg^{+2} co-factor to become bound in the active site (Lorimer, Badger and Andrews, 1976; Lorimer and Miziorko, 1980; Stec, 2012). Importantly, this carbamylated CO_2 molecule is not consumed in the catalytic reaction but is essential to enable proper binding of the Mg^{+2} co-factor as well as the RuBP substrate (Stec, 2012). In contrast, Mg^{+2} is essential for the catalytic chemistry of the rubisco reaction and its binding thereby completes the activation process (Lorimer, Badger and Andrews, 1976; Lorimer and Miziorko, 1980). However, this activation is not permanent, and deactivation by reversible decarbamylation does occur (Sharkey, Badger and Andrews, 2001). Thus, rubisco requires constant activation priming through post-translational modifications.

Finally, in addition to its competing reaction with O_2, Form I rubisco is also subject to competitive inhibition by numerous other molecules that result in enzyme inactivation. The predominant source of this inactivation arises because of improper binding of the RuBP substrate prior to active site carbamylation (Price *et al.*, 1995). However, aside from substrate auto-inhibition, rubisco in both active and inactive states can also be inhibited by diverse sugar phosphate molecules that are structurally analogous to RuBP (Orr *et al.*, 2023). Specifically, these ligands either represent misfire products of the rubisco reaction (e.g., xylulose 1,5-bisphosphate, D-Glycero-2,3-pentodiulose 1,5-bisphosphate and carboxy-tetritol-1,5-bisphosphate) or in the case of 2-carboxy-D-arabinitol 1-phosphate (CA1P) are synthesized to protect against nocturnal degradation (Moore *et al.*, 1992; Khan *et al.*, 1999; Parry *et al.*, 2008; Andralojc *et al.*, 2012). Moreover, because binding of RuBP as well as these substrate mimics is potent, the active site also subsequently requires functional repair

with further costs of ATP. This involves an initial step mediated by an ATPase Associated Activity (AAA+) rubisco activase chaperone (Mueller-Cajar *et al.*, 2011; Tsai *et al.*, 2015) to first remove the sugar phosphate (Robinson and Portis, 1988; Holbrook, Galasinski and Salvucci, 1991; Mueller-Cajar, Stotz and Bracher, 2014) as well as a secondary step mediated by a sugar phosphatase to render this molecule non-inhibitory (Holbrook, Bowes and Salvucci, 1989; Moore, Kobza and Seemann, 1991; Moore and Seemann, 1992). Thus, rubisco is subject to metabolic regulation and repair that constrains its CO_2 assimilation capacity.

1.3.5 The rubisco paradox: a dichotomy between an important, yet inefficient, enzyme

Given its foundational importance to life on our planet, the inherent inefficiencies of rubisco are puzzling. Specifically, the physical and chemical limitations of the enzyme place rubisco at the centre of a paradox that is challenging to reconcile with the fundamental principles of Darwinian evolution. Foremost among this confusion is the central question: why is this enzyme which underpins nearly all life on Earth an inefficient catalyst under present conditions? Curiously, this question becomes even more perplexing when considering that evolution has routinely overcome rubisco's catalytic shortcomings through costly and complex solutions that either involve investing in enormous quantities of the enzyme (e.g., representing up to 50% total soluble protein and 25% total nitrogen in C_3 plants (Carmo-Silva *et al.*, 2015)), or by operating CO_2-concentrating mechanisms (CCMs). Together, this evolutionary enigma – which we here coin the "rubisco paradox" – has motivated a great deal of research interest into rubisco that has been sustained since the enzyme was first discovered as fraction 1 protein in the mid-20th century (Sharkey, 2023).

1.3.6 The rubisco paradox has important implications for crop engineering

Understanding the mechanistic basis of the rubisco paradox is fundamentally interesting to resolve the mystery which shrouds the evolution of (arguably) the world's most important enzyme, and by extension, that of life on Earth itself. Unravelling this paradox will also prove invaluable from an applied context of crop improvement. This is because rubisco represents a key rate-limiting step in CO_2 assimilation under steady-state conditions (Farquhar, von Caemmerer and Berry, 1980). More precisely, the seminal Farquhar-von Caemmerer-Berry "FvCB" biochemical model of photosynthesis stipulates that the assimilation capacity of photosynthesis is dependent upon the catalytic efficiency

of rubisco under conditions of ambient or sub-ambient CO_2 concentrations and saturating light (Farquhar, von Caemmerer and Berry, 1980). Predicated upon this, it follows that improving rubisco should offer a promising strategy by which to enhance the productivity of our global food, fibre, and fuel crops (Bainbridge *et al.*, 1995; Carmo-Silva *et al.*, 2015; Sharwood, 2017; Bailey-Serres *et al.*, 2019; Iñiguez, Aguiló-Nicolau and Galmés, 2021; Croce *et al.*, 2024). However, it should be stressed that this ambitious goal will first require a comprehensive appreciation of the rubisco paradox. This is because this enigma will provide key answers to address whether rubisco can even be improved beyond its current state in nature, and if it can, the scope of any such improvement that might be expected. Importantly, resolving the rubisco paradox may also help inform optimal downstream engineering strategies for improving the enzyme's catalysis. Together, unravelling the biological basis of the rubisco paradox thus holds the keys to understanding the past, present, and future potential of rubisco in both nature as well as in synthetic engineering contexts.

1.4 The form and function of stomata in plant photosynthesis

1.4.1 Physical constraints and the challenges of CO_2 diffusion in plants

Although the primary biochemical constraint on photosynthesis is the carboxylation efficiency of rubisco (Farquhar, von Caemmerer and Berry, 1980), in plants the rate of rubisco-mediated carbon fixation is also limited by unique physical constraints on CO_2 diffusion (Flexas *et al.*, 2013). This is because in contrast to their primitive green algal-like ancestors, the diversification of the land plant lineage was associated with an increase in morphological complexity and the differentiation of distinct tissue types. Notably, this transition led to the formation of leaves to house the specialised photosynthetic mesophyll cells of the plant (Harrison and Morris, 2018). Importantly, it also led to the innovation of an impermeable waxy cuticle to protect the epidermal surface of aerial tissues (including leaves) from desiccation (Yeats and Rose, 2013). In this way, it follows that the transition of photosynthesis onto land created new resistances and barriers that together impede the supply of the CO_2 substrate for photosynthesis (Evans *et al.*, 2009). To help overcome a key obstacle along this diffusion path, land plants evolved stomata to mediate CO_2 uptake into the leaf.

1.4.2 Stomata control the exchange of CO_2 and water vapor in plants

Stomata are microscopic pores which exchange CO_2 and water vapour between plants and their environment. Specifically, atmospheric CO_2 enters the leaf through stomata to provide the primary substrate for photosynthesis. Simultaneously, water vapour evaporates from the stomata along an opposing concentration gradient to CO_2 through the plant transpiration stream. Although foliar transpiration is crucial to plant survival and underpins the regulation of diverse physiological processes, it can lead to desiccation, hydraulic failure, or even death if water loss exceeds soil water uptake (Nilson and Assmann, 2007). Thus, the process of photosynthesis at the basis of plant growth and development functions at the expense of dehydration through transpiration. Together, this antagonism places the stomata at the centre of a fundamental trade-off in biology which has important consequences for plant productivity.

Given their central role in gas exchange, it is evident that stomata represent a key adaptation that underpinned the plant colonisation of land (Edwards, Kerp and Hass, 1998; Raven, 2002; Clark *et al.*, 2022). This is because in tandem with the waxy cuticle, the evolution of stomata restricted the diffusion of all gaseous fluxes to a single interface. The significance of the innovation of stomata, as well as the fundamental compromise between carbon uptake and water loss that they balance, is underscored by both the lethality of mutants which are devoid of stomata (MacAlister, Ohashi-Ito and Bergmann, 2007) as well as the reduction in fitness of mutants that produce stomata in excessive quantities (Pillitteri *et al.*, 2007). It thus follows that stomata are significant components of global primary production and have had a key role in shaping our Earth system processes, with 40% of all atmospheric CO_2 (Ciais *et al.*, 1997) and 70% of our water consumption (Blatt, Chaumont and Farquhar, 2014) diffusing across these pores on an annual basis.

1.4.3 Stomata dynamically modulate gas exchange in response to changing environments

Although stomata represent a key adaptation that facilitated the evolution of oxygenic photosynthesis on land, this transition was nonetheless associated with an unprecedented suite of new challenges. Primarily, this is because terrestrial habitats present heterogeneous and dynamic environments that can drive rapid perturbations in the rates of underlying biological processes (Ferroni and Živčak,

2023). Moreover, for the first time in the history of oxygenic photosynthesis, water became a limiting resource with notable consequences in determining the structure of terrestrial ecosystems (Craine *et al.*, 2012). Thus, in order to survive, plants were required to modulate the rates of carbon uptake and water loss in response to changes in both physiological demands as well as their external conditions. As a means to address this, stomata have evolved complex functionality over the course of land plant evolution that is emphasized by their specialized structure in the present-day.

Each individual stomate consists of a pair of guard cells that surround a central opening to form an elliptical pore. In this arrangement, the stomata function like a hydromechanical valve which can either open or close to alter the exchange rate of both CO_2 and water vapour. These reversible adjustments in stomatal aperture are modulated by guard cell turgor pressure, which is in turn driven by changes in the metabolism and transport of inorganic ions (K^+ and Cl^-) and organic solutes (malate and sucrose) (Lloyd, 1908; Allaway and Hsiao, 1973; Schnabl and Raschke, 1980; Outlaw, 1983; Tallman and Zeiger, 1988; Reckmann, Scheibe and Raschke, 1990; Poffenroth, Green and Tallman, 1992; Talbott and Zeiger, 1993, 1996; Azoulay-Shemer *et al.*, 2015). In brief, an increase in guard cell osmotic content drives water influx into these cells which causes them to swell and the stomatal pore to open. In contrast, a reduction in guard cell osmotic content causes cell shrinking and closure of the central stomatal pore. Through this mechanism, stomata thus regulate the diffusive conductance of the plant by modulating the aperture of the pore through which all gaseous exchange occurs.

As a core adaptation of sessile plants to life in fluctuating terrestrial habitats, these dynamic stomatal movements are sensitive to both environmental and endogenous stimuli over time and space. Specifically, stomatal behaviour is co-ordinated by a myriad of signal transduction pathways which are responsive to changes in light, humidity, temperature, soil water availability, and atmospheric CO_2 (MacRobbie and Lettau, 1980; Amodeo, Talbott and Zeiger, 1996; Talbott and Zeiger, 1996; Schroeder *et al.*, 2001; Kim *et al.*, 2010; Araújo, Fernie and Nunes-Nesi, 2011; Hauser, Waadt and Schroeder, 2011; Daszkowska-Golec and Szarejko, 2013; Lawson and Matthews, 2020). Generally, these regulatory mechanisms enable stomata to open during conditions which are favourable for CO_2 fixation, and to subsequently close at night, or when water uptake is insufficient to keep pace

with evaporative demand. In this way, it follows that stomata do not represent passive interfaces for plant gas exchange. Instead, they dynamically fine-tune the balance of plant carbon gain and water loss to optimise plant performance over a diverse range of conditions (Cowan and Farquhar, 1977).

1.4.4 Stomata modulate gas exchange across longer periods of time

Although stomata regulate plant diffusive conductance in response to the environment over dynamic time scales, these movements are an order of magnitude slower than the associated changes that occur in photosynthesis (Lawson and Blatt, 2014; McAusland *et al.*, 2016; Lawson and Vialet-Chabrand, 2019). Accordingly, the discord between the respective rates of changes in photosynthesis and transpiration present a significant limitation on productivity which can equate to substantial costs in unrealised carbon potential and excessive water loss (Lawson and Blatt, 2014; McAusland *et al.*, 2016; Lawson and Vialet-Chabrand, 2019). To combat this, stomata also optimise their gas exchange properties through long-term adaptations to their growth conditions through changes in their physical properties (Qi and Torii, 2018; Driesen *et al.*, 2020).

Changes in guard cell morphology represent one important stomatal adaptation which has occurred over the course of land plant evolution. Specifically, whilst kidney-shaped guard cells represent the ancestral state found in most plants, grasses and some monocot lineages have evolved dumbbell-shaped guard cells which enable faster turgor-driven aperture responses owing to their larger surface area to volume ratio for changes in osmotic pressure (Franks and Farquhar, 2007; McAusland *et al.*, 2016). Moreover, these guard cells are also associated with the novel specialisation of flanking subsidiary cells that also exhibit turgor-driven changes and thus provide an additional layer of control on stomatal movements (Franks and Farquhar, 2007). Together, these derived adaptations in stomatal form and function have thus enabled specific plant lineages to achieve tighter co-ordination between photosynthesis and transpiration.

Beyond species-specific adaptations, stomata also demonstrate considerable developmental plasticity to changes in prevailing growth conditions. For instance, the distribution of stomata between leaf abaxial and adaxial surfaces influences the diffusion path length of CO_2 and H_2O gases (Xiong and Flexas, 2020) and varies in some species with changes in light intensity (James and Bell, 2000). Furthermore, both stomatal density and size are key anatomical traits which scale with leaf

maximal conductance and exhibit remarkable flexibility (Doheny-Adams *et al.*, 2012; de Boer *et al.*, 2016; Dittberner *et al.*, 2018; Buckley, Caine and Gray, 2020). Of these, stomatal density is the most important anatomical constraint on stomatal conductance and can be modulated in newly developing leaves in response to changes in light (Vigna *et al.*, 1980), CO_2 concentration (Woodward and Bazzaz, 1988; Woodward, Lake and Quick, 2002), relative air humidity (Bakker, 1991; Fanourakis, Heuvelink and Carvalho, 2013), soil water availability (Quarrie and Jones, 1977) and temperature (Lau *et al.*, 2018). In this way, stomatal densities have also adapted to changes in atmospheric CO_2 concentrations over both recent and geological time scales (Woodward, 1987; McElwain and Chaloner, 1995; Beerling, McElwain and Osborne, 1998; Beerling, Osborne and Chaloner, 2001; Franks and Beerling, 2009), highlighting their key importance as a plant adaptive strategy. Taken together with dynamic stomatal movements, developmental changes in the form and function of stomata thus represent an important mechanism by which plants optimise their physiology across diverse terrestrial environments and ecological contexts.

1.4.5 Stomata present a key target for engineering enhanced plant performance

Given their critical role as gatekeepers of plant gas exchange, it follows that manipulating stomata has represented a key target of crop improvement for over four decades (Farquhar and Richards, 1984). Specifically, stomatal breeding and engineering efforts have long sought to optimise the photosynthesis-water relations of crops (Bertolino, Caine and Gray, 2019; Endo and Torii, 2019; Faralli, Matthews and Lawson, 2019; Leakey *et al.*, 2019; Nguyen *et al.*, 2023). Central to this goal is the improvement of plant water-use efficiency (WUE) which provides a useful assessment of plant productivity by measuring the relative rates of the competing processes of CO_2 assimilation and transpiration. More precisely, the intrinsic WUE (*i*WUE) of the plant is calculated from the net CO_2 assimilation (A) that is measured in a fixed area of leaf divided by the stomatal conductance to water vapour (g_s) at an instantaneous moment in time under a given set of environmental conditions.

According to the respective parameters and the method in which *i*WUE is calculated (i.e., *i*WUE = A/g_s), it follows that optimising this trait can be achieved through two disparate strategies under a fixed set of environmental conditions. Specifically, this includes either increasing the relative net rate of leaf-level CO_2 assimilation or reducing the relative rate of leaf stomatal conductance to water

vapour. However, of these two respective approaches, stomatal conductance is a more important determinant of plant WUE (McAusland *et al.*, 2016) owing to the considerably larger size of this flux (routinely measured in units of mol m^{-2} s^{-1}) compared to the equivalent flux of CO_2 assimilation (routinely measured in units of μmol m^{-2} s^{-1}). Accordingly, engineering stomata provides a more promising avenue for enhancing crop WUE in our agricultural systems compared to improving the inherent efficiency of photosynthesis.

In light of its potential as a bioengineering handle, manipulating stomata form and function has attracted considerable interest from disparate engineering efforts. Among these diverse strategies which have been adopted, stomatal density has emerged as a key target of paramount interest given both its developmental plasticity as well as its central role in the adaptation of plants in nature (Hetherington and Woodward, 2003; Bertolino, Caine and Gray, 2019; Leakey *et al.*, 2019). Accordingly, an extensive number of stomatal density mutants have been previously generated across a wide range of studies (Yang and Sack, 1995; Schlüter *et al.*, 2003; Bergmann, Lukowitz and Somerville, 2004; Masle, Gilmore and Farquhar, 2005; Bussis *et al.*, 2006; Pillitteri *et al.*, 2007; Hara *et al.*, 2007, 2009; Yoo *et al.*, 2010; Doheny-Adams *et al.*, 2012; Gudesblat *et al.*, 2012; Tanaka *et al.*, 2013; Hepworth *et al.*, 2015; Franks *et al.*, 2015; Wang *et al.*, 2016; Hughes *et al.*, 2017; Yin *et al.*, 2017; Jalakas *et al.*, 2018; Schuler *et al.*, 2018; Mohammed *et al.*, 2019; Caine *et al.*, 2019; Dunn *et al.*, 2019; Sakoda *et al.*, 2020; Harrison *et al.*, 2020; Nguyen *et al.*, 2023; Zhao *et al.*, 2024).

A general pattern across engineered plant mutants is that manipulation of leaf stomatal density positively correlates with leaf steady-state stomatal conductance to both CO_2 and H_2O. As such, mutants that have been engineered to exhibit increased stomatal densities compared to unmodified wild-type control display enhanced steady-state conductance to both gases, and *vice versa*. However, the changes in the downstream rates of biochemical CO_2 assimilation and transpiration are not affected to equivalent extents. Specifically, this discord in the rates of these processes occurs because of differences in the partial pressures and diffusion constants of both CO_2 and H_2O gases, as well as the nonlinear response of photosynthesis to CO_2 (Farquhar, von Caemmerer and Berry, 1980). Thus, increasing stomatal density often results in an enhancement of leaf-level photosynthesis but a simultaneous deterioration of overall plant WUE. Inversely, engineering

reduced stomatal density in plants can often drive an enhancement of WUE such that more CO_2 is fixed per unit water transpired by the leaf. Together, engineering stomata thus provides a more attractive target compared to photosynthesis to improve crop WUE. Moreover, these improvements in plant WUE should be achieved by efforts seeking to engineer reductions, rather than increases, in stomatal density.

1.4.6 Stomatal development: knowledge gaps, outstanding questions, and future opportunities

The vast majority of all previous stomatal density engineering efforts have relied on perturbing the expression levels of endogenous developmental genes. Accordingly, this work has built upon decades of basic research which has sought to elucidate the underlying genetic networks that co-ordinate stomatal development. Important discoveries include the identification of the three master transcription factors SPEECHLESS (SPCH), MUTE and FAMA which mediate the entry, proliferation, and differentiation of the stomatal lineage (Bergmann, Lukowitz and Somerville, 2004; Pillitteri *et al.*, 2007; Gudesblat *et al.*, 2012). Significant discoveries also include the identification of multiple extracellular ligand-receptor components including the EPIDERMAL PATTERNING FACTOR (EPF)/EPF-like (EPFL) family, STOMATAL DENSITY AND DISTRIBUTION 1 (SDD1), TOO MANY MOUTHS (TMM), and the ERECTA kinases which together interact with these master regulators through signalling cascades to help co-ordinate stomatal patterning (Casson and Gray, 2008). As such, it follows that considerable advancements in mapping the genetic networks of stomatal development have been made and have offered multiple targets for crop improvement.

Despite the above successes in elucidating the genetic components of stomatal development, little is currently known about how transcriptional or functional changes in these genes and proteins are in turn modulated in response to metabolism or fluctuations in growth conditions (Hetherington and Woodward, 2003). This represents an important knowledge gap in our understanding of both plant development and plant adaptation that has far-reaching implications in stomatal biology, physiology, ecology, evolution, and global change. Importantly, this understanding also holds considerable value in applied contexts as would offer new handles that can be leveraged for manipulating stomatal density in crop improvement. Because plant metabolism is highly responsive to environmental

change, it is possible that metabolic changes in the leaf present one missing component in the web of control systems that regulate stomatal lineage development. Accordingly, unravelling the metabolic control of stomatal development presents an important goal of both fundamental research as well as ongoing efforts to engineer more productive and water-efficient crops.

1.5 Thesis overview

1.5.1 Thesis aims and objectives

The overarching goal of this thesis is to advance our understanding of the evolution of carbon fixation in plants and to explore the scope of optimising this process through genetic engineering. To achieve this feat, this research harnesses diverse tools in the current computational and synthetic biology revolutions to answer important questions concerning how carbon fixation functions, how it has been altered through time, and how it can be further adapted in future. In brief, this thesis explores these collective questions across two conceptually distinct, but complementary research streams which are focused on rubisco (Aims #1, #2, #3 and #4) and stomata (Aim #5), respectively. More precisely, this body of work can be distilled into the following set of five aims.

Aim 1

- To evaluate the extent by which evolutionary and biochemical factors constrain rubisco kinetic adaptation. This aim is explored in **Chapter 2** and **Appendix A**, respectively.

Aim 2

- To quantify the pace of rubisco molecular evolution through time to provide insight on the dynamics of enzyme adaptation. This aim is explored in **Chapter 3**.

Aim 3

- To assess whether rubisco catalysis is in evolutionary stasis, or whether the enzyme is adaptively evolving for improved fitness. This aim is explored in **Chapter 3**.

Aim 4

- To leverage the results generated from Aims 1 – 3 to provide additional mechanistic insight on the rubisco paradox, with potential implications for enzyme engineering. This aim is explored as the subject of a dedicated discussion and perspective in **Chapter 5**.

Aim 5

- To investigate whether leaf metabolism can be targeted to manipulate stomatal patterning, with potential implications for increasing plant productivity. This aim is explored in **Chapter 4**.

1.5.2 Thesis layout

Each data chapter of this thesis has been designed to represent a self-contained body of work that addresses one or more of the central aims of this thesis, as previously outlined. Accordingly, each of these chapter embodies a distinct publishable output which contains its own abstract, introduction, results, discussion, and method sections. The research presented in Chapter 2 (Bouvier *et al.*, 2021), Chapter 3 (Bouvier, Emms and Kelly, 2024), and Appendix A (Bouvier and Kelly, 2023) have already been published as peer-reviewed journal articles. Moreover, Chapter 4 (Bouvier and Kelly, 2024) is currently under journal review and is available as an online pre-print at the time of writing this thesis.

In addition to the above data chapters, this thesis also contains a separate general introduction (including literature review) and general discussion in order to formulate a coherent and continuous body of work that is unified around the central theme of: *The evolution and engineering of carbon fixation in plants*. To summarise, this body of work thus contains three data chapters, one additional data chapter presented as a supplemental appendix (Appendix A), and two distinct framing or contextualising chapters to build thesis narrative. Supplemental results and text associated with data Chapters 2 – 4 are also included as supplemental appendices (Appendices B – D, respectively). A detailed chapter-by-chapter outline is provided below to describe the key findings that are presented (primary data chapters) or the additional context that is provided (non-data chapters) in each respective section, with a view to demonstrate how these relate to the central theme and/or aims herein.

Chapter 1 – General Introduction

- *This chapter outlines the importance of crop improvement for our global humanity and highlights the current challenges which threaten our agricultural systems. Subsequently, this chapter introduces the opportunities for engineering carbon fixation to achieve food security in the 21st Century, with a key focus on rubisco and stomata as important determinants of plant photosynthetic efficiency.*

Chapter 2 – Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off

- *This data chapter presents a phylogenetically resolved analysis of rubisco kinetic adaptation. This demonstrates that rubisco kinetic traits have been evolving more independently of one other than previously assumed. This also discovers that evolutionary constraints explain more variation in rubisco kinetics compared to biochemical constraints underpinned by kinetic trait trade-offs.*

This thesis chapter has been published at the following location: Bouvier, J.W., Emms, D.M., Rhodes, T., Bolton, J.S., Brasnett, A., Eddershaw, A., Nielsen, J.R., Unitt, A., Whitney, S.M. and Kelly, S. 2021. *Molecular Biology and Evolution*, 38(7), 2880-2896

Chapter 3 – Rubisco is evolving for improved catalytic efficiency and CO_2 assimilation in plants

- *This data chapter presents an integrated phylogenetic interrogation of rubisco molecular and kinetic evolution. This finds that the rubisco catalytic large subunit (RbcL) is one of the slowest evolving sequences in all photosynthesis organisms that encode the Form I enzyme. This also demonstrates that rubisco is not in evolution stasis, but is instead continually evolving for improved catalysis and leaf-level photosynthesis.*

This thesis chapter has been published at the following location: Bouvier, J.W., Emms, D.M. and Kelly, S. 2024. *Proceedings of the National Academy of Sciences*, 121(11), Article No e2321050121.

Chapter 4 – Metabolic engineering of stomatal precursor cells enhances photosynthetic water-use efficiency and vegetative growth under water-deficit conditions in Arabidopsis thaliana

- *This data chapter presents a novel proof-of-concept metabolic approach to manipulate stomatal patterning in Arabidopsis thaliana. This demonstrates that NAD(P)$^+$/NAD(P)H ratios in stomatal precursor cells can be genetically engineered to reduce leaf stomatal density in plants. This also discovers that engineered plants exhibit improved photosynthetic water-use efficiency and display enhanced vegetative growth under water-deficit conditions.*

This thesis chapter has been published as pre-print at the following location: Bouvier, J.W. and Kelly, S. 2024. *bioRxiv*, doi 10.1101/2024.08.29.610053

Chapter 5 – General Discussion

- *This chapter provides an appraisal of the key findings in this thesis and how they related to the core aims herein. In accordance with the overarching structure of this research related to 'evolution', 'engineering', 'carbon fixation', 'rubisco', and 'stomata', a perspective of the wider implications of these findings in relation to these themes are then discussed in dedicated sections. In the specific context of Aim 4, this chapter argues that phylogenetic constraints (likely underpinned by molecular constraints on functional evolvability) provide a previously unidentified limitation on rubisco adaptation that offers new perspective to understand the "rubisco paradox".*

Appendix A – Response to Tcherkez and Farquhar: Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off

- *This data chapter presents a formal response to an Opinion article relating to Chapter 2 (published in: Tcherkez, G. and Farquhar, G.D. 2021. Journal of Plant Physiology, 267, Article No. 153554). This extended set of analyses confirm the work presented in Chapter 2 is valid and robust.*

This thesis chapter has been published at the following location: Bouvier, J.W. and Kelly, S. 2023. *Journal of Plant Physiology*, Article No 154021

1.5.3 Collaboration

Scientific progress thrives with collaboration. Whilst the research contained in this thesis has been driven predominantly by myself, I have not achieved this in isolation. It is important to acknowledge that this work has profited from the expert guidance, intellect, and technical knowledge of many colleagues. Accordingly, a comprehensive list of these collaborators, as well as the significance of their respective contributions, are outlined at the point at which each chapter commences. Where published, this information can also be found on the online version of these manuscripts.

1.6 References

Aigner, H. et al. (2017) 'Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2', Science, 358(6368), pp. 1272–1278. Available at: https://doi.org/10.1126/science.aap9221.

Ainsworth, E.A. and Long, S.P. (2005) 'What have we learned from 15 years of free-air CO2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO2', New Phytologist, 165(2), pp. 351–372. Available at: https://doi.org/10.1111/j.1469-8137.2004.01224.x.

Ainsworth, E.A. and Long, S.P. (2021) '30 years of free-air carbon dioxide enrichment (FACE): What have we learned about future crop productivity and its potential for adaptation?', Global Change Biology, 27(1), pp. 27–49. Available at: https://doi.org/10.1111/gcb.15375.

Al-Ghussain, L. (2019) 'Global warming: review on driving forces and mitigation', Environmental Progress and Sustainable Energy, 38(1), pp. 13–21. Available at: https://doi.org/10.1002/ep.13041.

Albery, W.J. and Knowles, J.R. (1976) 'Evolution of Enzyme Function and the Development of Catalytic Efficiency', Biochemistry, 15(25), pp. 5631–5640. Available at: https://doi.org/10.1021/bi00670a032.

Allaway, W.G. and Hsiao, T.C. (1973) 'Preparation of rolled epidermis of vicia faba l. So that stomata are the only viable cells: Analysis of guard cell potassium by flame photometry', Australian Journal of Biological Sciences, 26(2), pp. 309–318. Available at: https://doi.org/10.1071/BI9730309.

Amodeo, G., Talbott, L.D. and Zeiger, E. (1996) 'Use of potassium and sucrose by onion guard cells during a daily cycle of osmoregulation', Plant and Cell Physiology, 37(5), pp. 575–579. Available at: https://doi.org/10.1093/oxfordjournals.pcp.a028983.

Anderson, L.E. (1971) 'Chloroplast and cytoplasmic enzymes II. Pea leaf triose phosphate isomerases', BBA - Enzymology, 235(1), pp. 237–244. Available at: https://doi.org/10.1016/0005-2744(71)90051-9.

Anderson, L.E., Price, G.B. and Fuller, R.C. (1968) 'Molecular Diversity of the Ribulose-1,5-Diphosphate Carboxylase from Photosynthetic Microorganisms', Science, 161(3840), pp. 482–484. Available at: https://doi.org/10.1126/science.161.3840.482.

Andralojc, P.J. et al. (2012) '2-Carboxy-D-arabinitol 1-phosphate (CA1P) phosphatase: Evidence for a wider role in plant Rubisco regulation', Biochemical Journal, 442(3), pp. 733–742. Available at: https://doi.org/10.1042/BJ20111443.

Andrews, T.J. (1988) 'Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits.', The Journal of biological chemistry, 263(25), pp. 12213–12219. Available at: https://doi.org/10.1016/s0021-9258(18)37741-x.

Andrews, T.J., Greenwood, D.M. and Yellowlees, D. (1984) 'Catalytically active hybrids formed in vitro between large and small subunits of different procaryotic ribulose bisphosphate carboxylases', Archives of Biochemistry and Biophysics, 234(1), pp. 313–317. Available at: https://doi.org/10.1016/0003-9861(84)90354-0.

Andrews, T.J. and Lorimer, G.H. (1985) 'Catalytic properties of a hybrid between cyanobacterial large subunits and higher plant small subunits of ribulosebisphosphate carboxylase-oxygenase', Journal of Biological Chemistry, 260(8), pp. 4632–4636. Available at: https://doi.org/10.1016/s0021-9258(18)89117-7.

Aono, R. et al. (2015) 'A pentose bisphosphate pathway for nucleoside degradation in Archaea', Nature Chemical Biology, 11(5), pp. 355–360. Available at: https://doi.org/10.1038/nchembio.1786.

Araújo, W.L., Fernie, A.R. and Nunes-Nesi, A. (2011) 'Control of stomatal aperture: A renaissance of the old guard', Plant Signaling and Behavior, 6(9), pp. 1305–1311. Available at: https://doi.org/10.4161/psb.6.9.16425.

Archibald, J.M. (2015) 'Genomic perspectives on the birth and spread of plastids', Proceedings of the National Academy of Sciences of the United States of America, 112(33), pp. 10147–10153. Available at: https://doi.org/10.1073/pnas.1421374112.

Azoulay-Shemer, T. et al. (2015) 'Guard cell photosynthesis is critical for stomatal turgor production, yet does not directly mediate CO2- and ABA-induced stomatal closing', Plant Journal, 83(4), pp. 567–581. Available at: https://doi.org/10.1111/tpj.12916.

Bailey-Serres, J. et al. (2019) 'Genetic strategies for improving crop yields', Nature, 575(7781), pp. 109–118. Available at: https://doi.org/10.1038/s41586-019-1679-0.

Bainbridge, G. et al. (1995) 'Engineering Rubisco to change its catalytic properties', Journal of Experimental Botany, 46(special_issue), pp. 1269–1276. Available at: https://doi.org/10.1093/jxb/46.special_issue.1269.

Bakker, J.C. (1991) 'Effects of humidity on stomatal density and its relation to leaf conductance', Scientia Horticulturae, 48(3–4), pp. 205–212. Available at: https://doi.org/10.1016/0304-4238(91)90128-L.

Bar-Even, A. et al. (2011) 'The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters', Biochemistry, 50(21), pp. 4402–4410. Available at: https://doi.org/10.1021/bi2002289.

Bar-Even, A. et al. (2012) 'Thermodynamic constraints shape the structure of carbon fixation pathways', Biochimica et Biophysica Acta - Bioenergetics, 1817(9), pp. 1646–1659. Available at: https://doi.org/10.1016/j.bbabio.2012.05.002.

Bar-On, Y.M. and Milo, R. (2019) 'The global mass and average rate of rubisco', Proceedings of the National Academy of Sciences of the United States of America, 116(10), pp. 4738–4743. Available at: https://doi.org/10.1073/pnas.1816654116.

Barber, J. (2004) 'Engine of life and big bang of evolution: A personal perspective', Photosynthesis Research, 80(1–3), pp. 137–155. Available at: https://doi.org/10.1023/B:PRES.0000030662.04618.27.

Bartholomay, A.F. (1962) 'A Stochastic Approach to Statistical Kinetics with Application to Enzyme Kinetics', Biochemistry, 1(2), pp. 223–230. Available at: https://doi.org/10.1021/bi00908a005.

Bassham, J.A. et al. (1954) 'The Path of Carbon in Photosynthesis. XXI. The Cyclic Regeneration of Carbon Dioxide Acceptor', Journal of the American Chemical Society, 76(7), pp. 1760–1770. Available at: https://doi.org/10.1021/ja01636a012.

Bathellier, C. et al. (2018) 'Rubisco is not really so bad', Plant Cell and Environment, pp. 705–716. Available at: https://doi.org/10.1111/pce.13149.

Beerling, D.J., McElwain, J.C. and Osborne, C.P. (1998) 'Stomatal responses of the "living fossil" Ginkgo biloba L. to changes in atmospheric CO2 concentrations', Journal of Experimental Botany, 49(326), pp. 1603–1607. Available at: https://doi.org/10.1093/jxb/49.326.1603.

Beerling, D.J., Osborne, C.P. and Chaloner, W.G. (2001) 'Evolution of leaf-form in land plants linked to atmospheric CO2 decline in the Late Palaeozoic era', Nature, 410(6826), pp. 352–354. Available at: https://doi.org/10.1038/35066546.

Bekker, A. et al. (2004) 'Dating the rise of atmospheric oxygen', Nature, 427(6970), pp. 117–120. Available at: https://doi.org/10.1038/nature02260.

Bergmann, D.C., Lukowitz, W. and Somerville, C.R. (2004) 'Stomatal development and pattern controlled by a MAPKK kinase', Science, 304(5676), pp. 1494–1497. Available at: https://doi.org/10.1126/science.1096014.

Bertolino, L.T., Caine, R.S. and Gray, J.E. (2019) 'Impact of stomatal density and morphology on water-use efficiency in a changing world', Frontiers in Plant Science, 10. Available at: https://doi.org/10.3389/fpls.2019.00225.

Betts, R.A. et al. (2024) 'Mauna Loa carbon dioxide forecast for 2024 : Atmospheric CO 2 rise predicted to exceed IPCC 1 . 5 ° C scenarios', pp. 1–11.

Blatt, M.R., Chaumont, F. and Farquhar, G. (2014) 'Focus on water', Plant Physiology, 164(4), pp. 1553–1555. Available at: https://doi.org/10.1104/pp.114.900484.

Bloom, A.J. and Lancaster, K.M. (2018) 'Manganese binding to Rubisco could drive a photorespiratory pathway that increases the energy efficiency of photosynthesis', Nature Plants, 4(7), pp. 414–422. Available at: https://doi.org/10.1038/s41477-018-0191-0.

de Boer, H.J. et al. (2016) 'Optimal allocation of leaf epidermal area for gas exchange', New Phytologist, 210(4), pp. 1219–1228. Available at: https://doi.org/10.1111/nph.13929.

Borlaug, N.E. (2003) 'Feeding A World of 10 Billion People: Our 21st Century Challenge', Perspectives in World Food and Agriculture 2004, (May 2003), pp. 31–56. Available at: https://doi.org/10.1002/9780470290187.ch2.

Bouvier, J.W. et al. (2021) 'Rubisco Adaptation Is More Limited by Phylogenetic Constraint Than by Catalytic Trade-off', Molecular Biology and Evolution, 38(7), pp. 2880–2896. Available at: https://doi.org/10.1093/molbev/msab079.

Bouvier, J.W., Emms, D.M. and Kelly, S. (2024) 'Rubisco is evolving for improved catalytic efficiency and CO2 assimilation in plants', Proceedings of the National Academy of Sciences of the United States of America, 121(11). Available at: https://doi.org/10.1073/pnas.2321050121.

Bouvier, J.W. and Kelly, S. (2023) 'Response to Tcherkez and Farquhar: Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off', Journal of Plant Physiology, 287, p. 154021. Available at: https://doi.org/10.1016/j.jplph.2023.154021.

Bouvier, J.W. and Kelly, S. (2024) 'Metabolic engineering of stomatal precursor cells enhances photosynthetic water-use efficiency and vegetative growth under water-deficit conditions in Arabidopsis thaliana', bioRxiv [Preprint]. Available at: https://doi.org/doi.org/10.1101/2024.08.29.610053.

Bowes, G., Ogren, W.L. and Hageman, R.H. (1971) Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase, Biochemical and Biophysical Research Communications. Available at: https://doi.org/10.1016/0006-291X(71)90475-X.

Bracher, A. et al. (2011) 'Crystal structure of a chaperone-bound assembly intermediate of form i Rubisco', Nature Structural and Molecular Biology, 18(8), pp. 875–880. Available at: https://doi.org/10.1038/nsmb.2090.

Bracher, A. et al. (2017) 'Biogenesis and Metabolic Maintenance of Rubisco', Annual Review of Plant Biology, 68, pp. 29–60. Available at: https://doi.org/10.1146/annurev-arplant-043015-111633.

Brutnell, T.P. et al. (1999) 'BUNDLE SHEATH DEFECTIVE2, a Novel Protein Required for Post-Translational Regulation of the rbcL Gene of Maize', The Plant Cell, 11(5), p. 849. Available at: https://doi.org/10.2307/3870819.

Buchanan, B.B. (1992) 'Carbon dioxide assimilation in oxygenic and anoxygenic photosynthesis', Photosynthesis Research, 33(2), pp. 147–162. Available at: https://doi.org/10.1007/BF00039177.

Buckley, C.R., Caine, R.S. and Gray, J.E. (2020) 'Pores for Thought: Can Genetic Manipulation of Stomatal Density Protect Future Rice Yields?', Frontiers in Plant Science, 10. Available at: https://doi.org/10.3389/fpls.2019.01783.

Busch, F.A. (2020) 'Photorespiration in the context of Rubisco biochemistry, CO2 diffusion and metabolism', Plant Journal, 101(4), pp. 919–939. Available at: https://doi.org/10.1111/tpj.14674.

Busch, F.A., Sage, R.F. and Farquhar, G.D. (2018) 'Plants increase CO2 uptake by assimilating nitrogen via the photorespiratory pathway', Nature Plants, 4(1), pp. 46–54. Available at: https://doi.org/10.1038/s41477-017-0065-x.

Bussis, D. et al. (2006) 'Stomatal aperture can compensate altered stomatal density in Arabidopsis thaliana at growth light conditions', Functional Plant Biology, 33, pp. 1037–1043.

Caine, R.S. et al. (2019) 'Rice with reduced stomatal density conserves water and has improved drought tolerance under future climate conditions', New Phytologist, 221(1), pp. 371–384. Available at: https://doi.org/10.1111/nph.15344.

Calfapietra, C. et al. (2003) 'Free-air CO2 enrichment (FACE) enhances biomass production in a short-rotation poplar plantation', Tree Physiology, 23(12), pp. 805–814. Available at: https://doi.org/10.1093/treephys/23.12.805.

Cardona, T. (2019) 'Thinking twice about the evolution of photosynthesis', Open Biology, 9(3). Available at: https://doi.org/10.1098/rsob.180246.

Cardona, T., Murray, J.W. and Rutherford, A.W. (2015) 'Origin and evolution of water oxidation before the last common ancestor of the cyanobacteria', Molecular Biology and Evolution, 32(5), pp. 1310–1328. Available at: https://doi.org/10.1093/molbev/msv024.

Carmo-Silva, E. et al. (2015) 'Optimizing Rubisco and its regulation for greater resource use efficiency', Plant, Cell and Environment, pp. 1817–1832. Available at: https://doi.org/10.1111/pce.12425.

Casson, S. and Gray, J.E. (2008) 'Influence of environmental factors on stomatal development', New Phytologist, 178(1), pp. 9–23. Available at: https://doi.org/10.1111/j.1469-8137.2007.02351.x.

Cavanagh, A.P., Slattery, R. and Kubien, D.S. (2022) 'Temperature-induced changes in Arabidopsis Rubisco activity and isoform expression', Journal of Experimental Botany [Preprint]. Available at: https://doi.org/10.1093/jxb/erac379.

Ciais, P. et al. (1997) 'A three-dimensional synthesis study of δ18O in atmospheric CO2 1. Surface fluxes', Journal of Geophysical Research Atmospheres, 102(5), pp. 5857–5872. Available at: https://doi.org/10.1029/96jd02360.

Clark, J.W. et al. (2022) 'The origin and evolution of stomata', Current Biology, 32(11), pp. R539–R553. Available at: https://doi.org/10.1016/j.cub.2022.04.040.

Cloud, P. (1973) 'Paleoecological significance of the banded iron-formation', Economic Geology, 68(7), pp. 1135–1143. Available at: https://doi.org/10.2113/gsecongeo.68.7.1135.

Conlan, B. et al. (2019) 'BSD2 is a Rubisco-specific assembly chaperone, forms intermediary hetero-oligomeric complexes, and is nonlimiting to growth in tobacco', Plant Cell and Environment, 42(4), pp. 1287–1301. Available at: https://doi.org/10.1111/pce.13473.

Copley, S.D. (2003) 'Enzymes with extra talents: Moonlighting functions and catalytic promiscuity', Current Opinion in Chemical Biology, 7(2), pp. 265–272. Available at: https://doi.org/10.1016/S1367-5931(03)00032-2.

Cotton, C.A., Edlich-Muth, C. and Bar-Even, A. (2018) 'Reinforcing carbon fixation: CO2 reduction replacing and supporting carboxylation', Current Opinion in Biotechnology, 49, pp. 49–56. Available at: https://doi.org/10.1016/j.copbio.2017.07.014.

Cowan, I.R. and Farquhar, G.D. (1977) 'Stomatal function in relation to leaf metabolism and environment.', in Symposia of the Society for Experimental Biology, pp. 471–505.

Craine, J.M. et al. (2012) 'Resource limitation, tolerance, and the future of ecological plant classification', Frontiers in Plant Science, 3(OCT). Available at: https://doi.org/10.3389/fpls.2012.00246.

Croce, R. et al. (2024) 'Perspectives on improving photosynthesis to increase crop yield', The Plant Cell [Preprint]. Available at: https://doi.org/10.1093/plcell/koae132.

Czaja, A.D. et al. (2013) 'Biological Fe oxidation controlled deposition of banded iron formation in the ca. 3770Ma Isua Supracrustal Belt (West Greenland)', Earth and Planetary Science Letters, 363, pp. 192–203. Available at: https://doi.org/10.1016/j.epsl.2012.12.025.

Dai, A., Zhao, T. and Chen, J. (2018) 'Climate Change and Drought: a Precipitation and Evaporation Perspective', Current Climate Change Reports, 4(3), pp. 301–312. Available at: https://doi.org/10.1007/s40641-018-0101-6.

Daszkowska-Golec, A. and Szarejko, I. (2013) 'Open or close the gate - Stomata action under the control of phytohormones in drought stress conditions', Frontiers in Plant Science, 4(MAY). Available at: https://doi.org/10.3389/fpls.2013.00138.

Delwiche, C.F. (1999) 'Tracing the thread of plastid diversity through the tapestry of life', American Naturalist, 154(4 SUPPL.). Available at: https://doi.org/10.2307/2463984.

DePaoli, H.C. et al. (2014) 'Synthetic biology as it relates to CAM photosynthesis: Challenges and opportunities', Journal of Experimental Botany, 65(13), pp. 3381–3393. Available at: https://doi.org/10.1093/jxb/eru038.

Dittberner, H. et al. (2018) 'Natural variation in stomata size contributes to the local adaptation of water-use efficiency in Arabidopsis thaliana', Molecular Ecology, 27(20), pp. 4052–4065. Available at: https://doi.org/10.1111/mec.14838.

Dodd, A.N. et al. (2002) 'Crassulacean acid metabolism: Plastic, fantastic', Journal of Experimental Botany, pp. 569–580. Available at: https://doi.org/10.1093/jexbot/53.369.569.

Doheny-Adams, T. et al. (2012) 'Genetic manipulation of stomatal density influences stomatal size, plant growth and tolerance to restricted water supply across a growth carbon dioxide gradient', Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1588), pp. 547–555. Available at: https://doi.org/10.1098/rstb.2011.0272.

Doron, L. et al. (2014) 'The BSD2 ortholog in Chlamydomonas reinhardtii is a polysome-associated chaperone that co-migrates on sucrose gradients with the rbcL transcript encoding the Rubisco large subunit', Plant Journal, 80(2), pp. 345–355. Available at: https://doi.org/10.1111/tpj.12638.

Driesen, E. et al. (2020) 'Influence of environmental factors light, co2, temperature, and relative humidity on stomatal opening and development: A review', Agronomy, 10(12). Available at: https://doi.org/10.3390/agronomy10121975.

Dunn, J. et al. (2019) 'Reduced stomatal density in bread wheat leads to increased water-use efficiency', Journal of Experimental Botany, 70(18), pp. 4737–4747. Available at: https://doi.org/10.1093/jxb/erz248.

Duysens, L.N.M., Amesz, J. and Kamp, B.M. (1961) 'Two photochemical systems in photosynthesis', Nature, 190(4775), pp. 510–511. Available at: https://doi.org/10.1038/190510a0.

Edwards, D., Kerp, H. and Hass, H. (1998) 'Stomata in early land plants: An anatomical and ecophysiological approach', Journal of Experimental Botany, 49(SPEC. ISS.), pp. 255–278. Available at: https://doi.org/10.1093/jxb/49.special_issue.255.

Edwards, E.J. (2019) 'Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis', New Phytologist, pp. 1742–1755. Available at: https://doi.org/10.1111/nph.15851.

Eigenbrode, J.L. and Freeman, K.H. (2006) 'Late Archean rise of aerobic microbial ecosystems', Proceedings of the National Academy of Sciences of the United States of America, 103(43), pp. 15759–15764. Available at: https://doi.org/10.1073/pnas.0607540103.

Eisenhut, M., Roell, M.S. and Weber, A.P.M. (2019) 'Mechanistic understanding of photorespiration paves the way to a new green revolution', New Phytologist, 223(4), pp. 1762–1769. Available at: https://doi.org/10.1111/nph.15872.

Ellis, R.J. (1979) The most abundant protein in the world, Trends in Biochemical Sciences. Available at: https://doi.org/10.1016/0968-0004(79)90212-3.

Endo, H. and Torii, K.U. (2019) 'Stomatal development and perspectives toward agricultural improvement', Cold Spring Harbor Perspectives in Biology, 11(5). Available at: https://doi.org/10.1101/cshperspect.a034660.

Erb, T.J. and Zarzycki, J. (2018) A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2 fixing enzyme, Current Opinion in Biotechnology. Available at: https://doi.org/10.1016/j.copbio.2017.07.017.

Ermakova, M. et al. (2020) 'On the road to C4 rice: advances and perspectives', Plant Journal, 101(4), pp. 940–950. Available at: https://doi.org/10.1111/tpj.14562.

Evans, J.R. et al. (2009) ‘Resistances along the CO2 diffusion pathway inside leaves’, Journal of Experimental Botany, 60(8), pp. 2235–2248. Available at: https://doi.org/10.1093/jxb/erp117.

Evans, L.T. (1997) ‘Adapting and improving crops: The endless task’, Philosophical Transactions of the Royal Society B: Biological Sciences, 352(1356), pp. 901–906. Available at: https://doi.org/10.1098/rstb.1997.0069.

Evans, L.T. and Dunstone, R.L. (1970) ‘Some physiological aspects of evolution in wheat’, Australian Journal of Biological Sciences, 23(4), pp. 725–741. Available at: https://doi.org/10.1071/BI9700725.

Evenson, R.E. and Gollin, D. (2003) ‘Assessing the Impact of the Green Revolution, 1960 to 2000’, Science, 300(5620), pp. 758–762.

Ezaki, S. et al. (1999) ‘Presence of a Structurally Novel Type Ribulose-bisphosphate Carboxylase/Oxygenase in the Hyperthermophilic Archaeon’, Molecular Biology, 274(8), pp. 5078–5082.

Fanourakis, D., Heuvelink, E. and Carvalho, S.M.P. (2013) ‘A comprehensive analysis of the physiological and anatomical components involved in higher water loss rates after leaf development at high humidity’, Journal of Plant Physiology, 170(10), pp. 890–898. Available at: https://doi.org/10.1016/j.jplph.2013.01.013.

FAO (2009) ‘Global agriculture towards 2050’, High Level Expert Forum-How to feed the world 2050, pp. 1–4. Available at: http://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf.

FAO (2022) ‘Food Security and Nutrition Around The World. The State of Food Security and Nutrition in the World 2021.’, The State of Food Security and Nutrition in the World 2022 [Preprint]. Available at: https://doi.org/10.4060/cc0639en.

FAO et al. (2024) The State of Food Security and Nutrition in the World 2024 – Financing to end hunger, food insecurity and malnutrition in all its forms, The State of Food Security and Nutrition in the World 2024. Rome, Italy. Available at: https://doi.org/10.4060/cd1254en.

FAOSTAT (2021) ‘New Food Balance Sheets’, Browse data [Preprint].

Faralli, M., Matthews, J. and Lawson, T. (2019) ‘Exploiting natural variation and genetic manipulation of stomatal conductance for crop improvement’, Current Opinion in Plant Biology, pp. 1–7. Available at: https://doi.org/10.1016/j.pbi.2019.01.003.

Farquhar, G.D., von Caemmerer, S. and Berry, J.A. (1980) ‘A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species’, Planta, 149(1), pp. 78–90. Available at: https://doi.org/10.1007/BF00386231.

Farquhar, G.D. and Richards, R.A. (1984) ‘Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes.’, Australian Journal of Plant Physiology, 11(6), pp. 539–552. Available at: https://doi.org/10.1071/PP9840539.

Farquhar James, Bao Huiming and Thiemens Mark (2000) ‘Atmospheric Influence of Earth’s Earliest Sulfur Cycle’, Science [Preprint]. Available at: https://www.science.org/doi/pdf/10.1126/science.289.5480.756?casa_token=zQXGeHQpUaUAAAAA:7_67kL1bGtPr_mdhQVUA97-MUlFEE-Zqffc_91InSg4hTCnEYGxDQWp1F0xadMPRQJKi_ttg1tEMWOAl.

Feiz, L. et al. (2012) ‘Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation Factor1 is required for holoenzyme assembly in maize’, Plant Cell, 24(8), pp. 3435–3446. Available at: https://doi.org/10.1105/tpc.112.102012.

Feiz, L. et al. (2014) ‘A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants’, Plant Journal, 80(5), pp. 862–869. Available at: https://doi.org/10.1111/tpj.12686.

Ferroni, L. and Živčak, M. (2023) ‘Photosynthesis under Environmental Fluctuations: A Challenge for Plants, a Challenge for Researchers’, Plants, 12(24). Available at: https://doi.org/10.3390/plants12244146.

Field, C.B. et al. (1998) ‘Primary production of the biosphere: Integrating terrestrial and oceanic components’, Science, 281(5374), pp. 237–240. Available at: https://doi.org/10.1126/science.281.5374.237.

Fischer, R.A. et al. (1998) ‘Wheat yield progress associated with higher stomatal conductance and photosynthetic rate, and cooler canopies’, Crop Science, 38(6), pp. 1467–1475. Available at: https://doi.org/10.2135/cropsci1998.0011183X003800060011x.

Fischer, W.W., Hemp, J. and Johnson, J.E. (2016) ‘Evolution of Oxygenic Photosynthesis’, Annual Review of Earth and Planetary Sciences, 44, pp. 647–683. Available at: https://doi.org/10.1146/annurev-earth-060313-054810.

Fitchen, J.H. et al. (1990) ‘Residues in three conserved regions of the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase are required for quaternary structure’, Proceedings of the National Academy of Sciences of the United States of America, 87(15), pp. 5768–5772. Available at: https://doi.org/10.1073/pnas.87.15.5768.

Flexas, J. et al. (2013) ‘Diffusional conductances to CO2 as a target for increasing photosynthesis and photosynthetic water-use efficiency’, Photosynthesis Research, 117(1–3), pp. 45–59. Available at: https://doi.org/10.1007/s11120-013-9844-z.

Franks, P.J. et al. (2015) ‘Increasing water-use efficiency directly through genetic manipulation of stomatal density’, New Phytologist, 207(1), pp. 188–195. Available at: https://doi.org/10.1111/nph.13347.

Franks, P.J. and Beerling, D.J. (2009) ‘Maximum leaf conductance driven by CO2 effects on stomatal size and density over geologic time’, Proceedings of the National Academy of Sciences of the United States of America, 106(25), pp. 10343–10347. Available at: https://doi.org/10.1073/pnas.0904209106.

Franks, P.J. and Farquhar, G.D. (2007) ‘The mechanical diversity of stomata and its significance in gas-exchange control’, Plant Physiology, 143(1), pp. 78–87. Available at: https://doi.org/10.1104/pp.106.089367.

Furbank, R., Kelly, S. and von Caemmerer, S. (2023) ‘Photosynthesis and food security: the evolving story of C4 rice’, Photosynthesis Research, 158(2), pp. 121–130. Available at: https://doi.org/10.1007/s11120-023-01014-0.

Genkov, T. and Spreitzer, R.J. (2009) ‘Highly conserved small subunit residues influence Rubisco large subunit catalysis’, Journal of Biological Chemistry, 284(44), pp. 30105–30112. Available at: https://doi.org/10.1074/jbc.M109.044081.

George, D.M., Vincent, A.S. and Mackey, H.R. (2020) ‘An overview of anoxygenic phototrophic bacteria and their applications in environmental biotechnology for sustainable Resource recovery’, Biotechnology Reports, 28. Available at: https://doi.org/10.1016/j.btre.2020.e00563.

Getzoff, T.P. et al. (1998) ‘Chimeric Arabidopsis thaliana Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Containing a Pea Small Subunit Protein Is Compromised in Carbamylation’, Plant Physiology, 116(2), pp. 695–702. Available at: https://doi.org/10.1104/pp.116.2.695.

Gielen, B. et al. (2005) 'Net carbon storage in a poplar plantation (POPFACE) after three years of free-air CO2 enrichment', Tree Physiology, 25(11), pp. 1399–1408. Available at: https://doi.org/10.1093/treephys/25.11.1399.

Gionfriddo, M., Zang, K. and Hayer-Hartl, M. (2023) 'The challenge of engineering Rubisco for improving photosynthesis', FEBS Letters, 597(13), pp. 1679–1680. Available at: https://doi.org/10.1002/1873-3468.14678.

Grabsztunowicz, M. et al. (2015) 'A reversible decrease in ribulose 1,5-bisphosphate carboxylase/oxygenase carboxylation activity caused by the aggregation of the enzyme's large subunit is triggered in response to the exposure of moderate irradiance-grown plants to low irradiance', Physiologia Plantarum, 154(4), pp. 591–608. Available at: https://doi.org/10.1111/ppl.12322.

Gudesblat, G.E. et al. (2012) 'SPEECHLESS integrates brassinosteroid and stomata signalling pathways', Nature Cell Biology, 14(5), pp. 548–554. Available at: https://doi.org/10.1038/ncb2471.

Guo, Q. et al. (2009) 'Reconstructing Earth's surface oxidation across the Archean-Proterozoic transition', Geology, 37(5), pp. 399–402. Available at: https://doi.org/10.1130/G25423A.1.

Hall, N.P. et al. (1987) 'Characteristics of a photorespiratory mutant of barley (Hordeum vulgare L.) deficient in phosphogly collate phosphatase', Photosynthesis Research, 11(1), pp. 89–96. Available at: https://doi.org/10.1007/BF00117676.

Hanada, S. (2016) 'Anoxygenic photosynthesis —A photochemical reaction that does not contribute to oxygen reproduction—', Microbes and Environments, 31(1), pp. 1–3. Available at: https://doi.org/10.1264/jsme2.ME3101rh.

Hara, K. et al. (2007) 'The secretory peptide gene EPF1 enforces the stomatal one-cell-spacing rule', Genes and Development, 21(14), pp. 1720–1725. Available at: https://doi.org/10.1101/gad.1550707.

Hara, K. et al. (2009) 'Epidermal cell density is autoregulated via a secretory peptide, EPIDERMAL PATTERNING FACTOR 2 in arabidopsis leaves', Plant and Cell Physiology, 50(6), pp. 1019–1031. Available at: https://doi.org/10.1093/pcp/pcp068.

Harrison, C.J. and Morris, J.L. (2018) 'The origin and early evolution of vascular plant shoots and leaves', Philosophical Transactions of the Royal Society B: Biological Sciences, 373(1739). Available at: https://doi.org/10.1098/rstb.2016.0496.

Harrison, E.L. et al. (2020) 'The influence of stomatal morphology and distribution on photosynthetic gas exchange', Plant Journal, 101(4), pp. 768–779. Available at: https://doi.org/10.1111/tpj.14560.

Hauser, F., Waadt, R. and Schroeder, J.I. (2011) 'Evolution of abscisic acid synthesis and signaling mechanisms', Current Biology, 21(9). Available at: https://doi.org/10.1016/j.cub.2011.03.015.

Hauser, T. et al. (2015) 'Structure and mechanism of the Rubisco-assembly chaperone Raf1', Nature Structural and Molecular Biology, 22(9), pp. 720–728. Available at: https://doi.org/10.1038/nsmb.3062.

Hayer-Hartl, M. (2017) 'From chaperonins to Rubisco assembly and metabolic repair', Protein Science, 26(12), pp. 2324–2333. Available at: https://doi.org/10.1002/pro.3309.

Hayer-Hartl, M. and Hartl, F.U. (2020) 'Chaperone Machineries of Rubisco – The Most Abundant Enzyme', Trends in Biochemical Sciences, 45(9), pp. 748–763. Available at: https://doi.org/10.1016/j.tibs.2020.05.001.

Hedden, P. (2003) 'The genes of the Green Revolution', Trends in Genetics, 19(1), pp. 5–9. Available at: https://doi.org/10.1016/S0168-9525(02)00009-4.

Hedges, S.B. et al. (2004) 'A molecular timescale of eukaryote evolution and the rise of complex multicellular life', BMC Evolutionary Biology, 4. Available at: https://doi.org/10.1186/1471-2148-4-2.

Hengsdijk, H. and Langeveld, J.W.. (2009) Yield trends and yield gap analysis of major crops in the world.

Hepworth, C. et al. (2015) 'Manipulating stomatal density enhances drought tolerance without deleterious effect on nutrient uptake', New Phytologist, 208(2), pp. 336–341. Available at: https://doi.org/10.1111/nph.13598.

Hetherington, A.M. and Woodward, F.I. (2003) 'The role of stomata in sensing and driving environmental change', Nature, 424(6951), pp. 901–908. Available at: https://doi.org/10.1038/nature01843.

Hoffman, P.F. (2013) 'The Great Oxidation and a Siderian snowball Earth: MIF-S based correlation of Paleoproterozoic glacial epochs', Chemical Geology, 362, pp. 143–156. Available at: https://doi.org/10.1016/j.chemgeo.2013.04.018.

Hofmann, H.J. (1976) 'Precambrian Microflora, Belcher Islands, Canada: Significance and Systematics', Journal of Paleontology, 50(6), pp. 1040–1073. Available at: http://www.jstor.org/stable/1303547.

Hohmann-Marriott, M.F. and Blankenship, R.E. (2011) 'Evolution of photosynthesis', Annual Review of Plant Biology, 62, pp. 515–548. Available at: https://doi.org/10.1146/annurev-arplant-042110-103811.

Holbrook, G.P., Bowes, G. and Salvucci, M.E. (1989) 'Degradation of 2-Carboxyarabinitol 1-Phosphate by a Specific Chloroplast Phosphatase', Plant Physiology, 90(2), pp. 673–678. Available at: https://doi.org/10.1104/pp.90.2.673.

Holbrook, G.P., Galasinski, S.C. and Salvucci, M.E. (1991) 'Regulation of 2-carboxyarabinitol 1-phosphatase', Plant Physiology, 97(3), pp. 894–899. Available at: https://doi.org/10.1104/pp.97.3.894.

Horken, K.M. and Tabita, F.R. (1999) 'The "green" form I ribulose 1,5-bisphosphate carboxylase/oxygenase from the nonsulfur purple bacterium Rhodobacter capsulatus', Journal of Bacteriology, 181(13), pp. 3935–3941. Available at: https://doi.org/10.1128/jb.181.13.3935-3941.1999.

Hossain, A. et al. (2020) 'Agricultural Land Degradation: Processes and Problems Undermining Future Food Security', Environment, Climate, Plant and Vegetation Growth, pp. 17–61. Available at: https://doi.org/10.1007/978-3-030-49732-3_2.

Huang, J. Bin et al. (2012) 'The science of global warming', Advances in Climate Change Research, 3(3), pp. 174–178. Available at: https://doi.org/10.3724/SP.J.1248.2012.00174.

Hughes, J. et al. (2017) 'Reducing stomatal density in barley improves drought tolerance without impacting on yield', Plant Physiology, 174(2), pp. 776–787. Available at: https://doi.org/10.1104/pp.16.01844.

Imhoff, J.F. (2017) 'Anoxygenic phototrophic bacteria from extreme environments', Modern Topics in the Phototrophic Prokaryotes: Environmental and Applied Aspects, pp. 427–480. Available at: https://doi.org/10.1007/978-3-319-46261-5_13.

Iñiguez, C., Aguiló-Nicolau, P. and Galmés, J. (2021) 'Improving photosynthesis through the enhancement of Rubisco carboxylation capacity', Biochemical Society Transactions, 49(5), pp. 2007–2019. Available at: https://doi.org/10.1042/BST20201056.

IPCC (2019) Special Report on Climate Change and Land, Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

IPCC (2023) 'AR6 Synthesis Report Climate Change 2023. https://www.ipcc.ch/report/ar6/syr/'.

Ishikawa, C. et al. (2011) 'Functional incorporation of sorghum small subunit increases the catalytic turnover rate of rubisco in transgenic rice', Plant Physiology, 156(3), pp. 1603–1611. Available at: https://doi.org/10.1104/pp.111.177030.

Jaffe, A.L. et al. (2019) 'Lateral Gene Transfer Shapes the Distribution of RuBisCO among Candidate Phyla Radiation Bacteria and DPANN Archaea', Molecular Biology and Evolution, 36(3), pp. 435–446. Available at: https://doi.org/10.1093/molbev/msy234.

Jalakas, P. et al. (2018) 'ABA-mediated regulation of stomatal density is OST1-independent', Plant Direct, 2(9). Available at: https://doi.org/10.1002/pld3.82.

James, S.A. and Bell, D.T. (2000) 'Influence of light availability on leaf structure and growth of two Eucalyptus globulus ssp. globulus provenances', Tree Physiology, 20(15), pp. 1007–1018. Available at: https://doi.org/10.1093/treephys/20.15.1007.

Jiang, X. et al. (2023) 'The role of photorespiration in plant immunity', Frontiers in Plant Science, 14. Available at: https://doi.org/10.3389/fpls.2023.1125945.

Johnson, J.E. et al. (2013) 'Manganese-oxidizing photosynthesis before the rise of cyanobacteria', Proceedings of the National Academy of Sciences of the United States of America, 110(28), pp. 11238–11243. Available at: https://doi.org/10.1073/pnas.1305530110.

Johnson, J.E. et al. (2014) 'O2 constraints from Paleoproterozoic detrital pyrite and uraninite', Bulletin of the Geological Society of America, 126(5–6), pp. 813–830. Available at: https://doi.org/10.1130/B30949.1.

Johnson, M.P. (2016) 'Photosynthesis', Essays in Biochemistry, 60(3), pp. 255–273. Available at: https://doi.org/10.1042/EBC20160016.

Jones, R. and Mearns, L. (2003) 'Assessing future climate risks', (5).

Joshi, J. et al. (2015) 'Role of small subunit in mediating assembly of red-type Form I Rubisco', Journal of Biological Chemistry, 290(2), pp. 1066–1074. Available at: https://doi.org/10.1074/jbc.M114.613091.

Kanevski, I. et al. (1999) 'Plastome engineering of ribulose-1,5-bisphosphate carboxylase/oxygenase in tobacco to form a sunflower large subunit and tobacco small subunit hybrid', Plant Physiology, 119(1), pp. 133–141. Available at: https://doi.org/10.1104/pp.119.1.133.

Karkehabadi, S. et al. (2005) 'Chimeric small subunits influence catalysis without causing global conformational changes in the crystal structure of ribulose-1,5-bisphosphate carboxylase/oxygenase', Biochemistry, 44(29), pp. 9851–9861. Available at: https://doi.org/10.1021/bi050537v.

Kasting, J.F., Toon, O.B. and Pollack, J.B. (1987) 'How climate evolved on the terrestrial planets.', Scientific American, 256(2), pp. 90–97. Available at: https://doi.org/10.1038/scientificamerican0288-90.

Keeling, P.J. (2010) 'The endosymbiotic origin, diversification and fate of plastids', Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1541), pp. 729–748. Available at: https://doi.org/10.1098/rstb.2009.0103.

Keleti, T. and Welch, G.R. (1984) 'The evolution of enzyme kinetic power.', The Biochemical journal, 223(2), pp. 299–303. Available at: https://doi.org/10.1042/bj2230299.

Kelly, G.J. and Latzko, E. (1976) 'Inhibition of spinach-leaf phosphofructokinase by 2-phosphoglycollate', FEBS Letters, 68(1), pp. 55–58. Available at: https://doi.org/10.1016/0014-5793(76)80403-6.

Kelly, S. (2022) 'The quest for more food', Science, 377(6604), pp. 370–371. Available at: https://doi.org/10.1126/science.add3882.

Khan, S. et al. (1999) '2'-carboxy-D-arabitinol 1-phosphate protects ribulose 1,5-bisphosphate carboxylase/oxygenase against proteolytic breakdown', European Journal of Biochemistry, 266(3), pp. 840–847. Available at: https://doi.org/10.1046/j.1432-1327.1999.00913.x.

Khush, G.S. (1999) 'Green revolution: preparing for the 21st century', Genome, 42(4), pp. 646–655. Available at: https://doi.org/10.1139/gen-42-4-646.

Kim, T.H. et al. (2010) 'Guard cell signal transduction network: Advances in understanding abscisic acid, CO2, and Ca2+ signaling', Annual Review of Plant Biology, 61, pp. 561–591. Available at: https://doi.org/10.1146/annurev-arplant-042809-112226.

Kirpich, P.Z., Haman, D.Z. and Styles, S.W. (1999) 'Problems of Irrigation in Developing Countries', Journal of Irrigation and Drainage Engineering, 125(1), pp. 1–6. Available at: https://doi.org/10.1061/(asce)0733-9437(1999)125:1(1).

Koch, L.G. and Britton, S.L. (2008) 'Aerobic metabolism underlies complexity and capacity', Journal of Physiology, 586(1), pp. 83–95. Available at: https://doi.org/10.1113/jphysiol.2007.144709.

Kolesinski, P. et al. (2014) 'Rubisco Accumulation Factor 1 from Thermosynechococcus elongatus participates in the final stages of ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in Escherichia coli cells and in vitro', FEBS Journal, 281(17), pp. 3920–3932. Available at: https://doi.org/10.1111/febs.12928.

Kookana, R.S. et al. (2020) 'Urbanisation and emerging economies: Issues and potential solutions for water and food security', Science of the Total Environment, 732. Available at: https://doi.org/10.1016/j.scitotenv.2020.139057.

Kornhuber, K. et al. (2023) 'Risks of synchronized low yields are underestimated in climate and crop model projections', Nature Communications, 14(1). Available at: https://doi.org/10.1038/s41467-023-38906-7.

Kostov, R. V., Small, C.L. and McFadden, B.A. (1997) 'Mutations in a sequence near the N-terminus of the small subunit alter the CO2/O2 specificity factor for ribulose bisphosphate carboxylase/oxygenase', Photosynthesis Research, 54(2), pp. 127–134. Available at: https://doi.org/10.1023/A:1005967106993.

Kruger, E.L. and Volin, J.C. (2006) 'Reexamining the empirical relation between plant growth and leaf photosynthesis', Functional Plant Biology, 33(5), pp. 421–429. Available at: https://doi.org/10.1071/FP05310.

Kumar, P. (2007) 'Green Revolution and its impact on environment', International Journal of Research in Humanities & Soc. Sciences, 5(3), pp. 54–57.

Kumar, S. et al. (2017) 'TimeTree: A Resource for Timelines, Timetrees, and Divergence Times', Molecular biology and evolution, 34(7), pp. 1812–1819. Available at: https://doi.org/10.1093/molbev/msx116.

Kump, L.R. (2008) 'The rise of atmospheric oxygen', Nature, 451(7176), pp. 277–278. Available at: https://doi.org/10.1038/nature06587.

Kweku, D. et al. (2018) 'Greenhouse Effect: Greenhouse Gases and Their Impact on Global Warming', Journal of Scientific Research and Reports, 17(6), pp. 1–9. Available at: https://doi.org/10.9734/jsrr/2017/39630.

Lau, O.S. et al. (2018) 'Direct Control of SPEECHLESS by PIF4 in the High-Temperature Response of Stomatal Development', Current Biology, 28(8), pp. 1273-1280.e3. Available at: https://doi.org/10.1016/j.cub.2018.02.054.

Lawson, T. and Blatt, M.R. (2014) 'Stomatal size, speed, and responsiveness impact on photosynthesis and water use efficiency', Plant Physiology, 164(4), pp. 1556–1570. Available at: https://doi.org/10.1104/pp.114.237107.

Lawson, T. and Matthews, J. (2020) 'Guard Cell Metabolism and Stomatal Function', Annual Review of Plant Biology, 71, pp. 273–302. Available at: https://doi.org/10.1146/annurev-arplant-050718-100251.

Lawson, T. and Vialet-Chabrand, S. (2019) 'Speedy stomata, photosynthesis and plant water use efficiency', New Phytologist, 221(1), pp. 93–98. Available at: https://doi.org/10.1111/nph.15330.

Leakey, A.D.B. et al. (2019) 'Water Use Efficiency as a Constraint and Target for Improving the Resilience and Productivity of C 3 and C 4 Crops', Annual Review of Plant Biology, 70, pp. 781–808. Available at: https://doi.org/10.1146/annurev-arplant-042817-040305.

Lee, B., Berka, R.M. and Tabita, F.R. (1991) 'Mutations in the small subunit of cyanobacterial ribulose-bisphosphate carboxylase/oxygenase that modulate interactions with large subunits', Journal of Biological Chemistry, 266(12), pp. 7417–7422. Available at: https://doi.org/10.1016/s0021-9258(20)89463-0.

Lee, B., Read, B.A. and Tabita, F.R. (1991) 'Catalytic properties of recombinant octameric, hexadecameric, and heterologous cyanobacterial/ bacterial ribulose-1,5-bisphosphate carboxylase/oxygenase', Archives of Biochemistry and Biophysics, 291(2), pp. 263–269. Available at: https://doi.org/10.1016/0003-9861(91)90133-4.

Lee, B. and Tabita, F.R. (1990) 'Purification of Recombinant Ribulose-l,5-bisphosphate Carboxylase/Oxygenase Large Subunits Suitable for Reconstitution and Assembly of Active L8S8 Enzyme', Biochemistry, 29(40), pp. 9352–9357. Available at: https://doi.org/10.1021/bi00492a007.

Lilley, J.M., Bolger, T.P. and Gifford, R.M. (2001) 'Productivity of Trifolium subterraneum and Phalaris aquatica under warmer, high CO2 conditions', New Phytologist, 150(2), pp. 371–383. Available at: https://doi.org/10.1046/j.1469-8137.2001.00108.x.

Lin, M.T. et al. (2020) 'Small subunits can determine enzyme kinetics of tobacco Rubisco expressed in Escherichia coli', Nature Plants, 6(10), pp. 1289–1299. Available at: https://doi.org/10.1038/s41477-020-00761-5.

Liu, C. et al. (2010) 'Coupled chaperone action in folding and assembly of hexadecameric Rubisco', Nature, 463(7278), pp. 197–202. Available at: https://doi.org/10.1038/nature08651.

Liu, X. et al. (2022) 'Global Agricultural Water Scarcity Assessment Incorporating Blue and Green Water Availability Under Future Climate Change', Earth's Future, 10(4). Available at: https://doi.org/10.1029/2021EF002567.

Lloyd, F.E. (1908) The physiology of Stomata, The physiology of Stomata. Carnegie Institution of Washington. Available at: https://doi.org/10.5962/bhl.title.24716.

Long, S.P. et al. (2006) ‘Can improvement in photosynthesis increase crop yields?’, Plant, Cell and Environment, 29(3), pp. 315–330. Available at: https://doi.org/10.1111/j.1365-3040.2005.01493.x.

Long, S.P., Marshall-Colon, A. and Zhu, X.G. (2015) ‘Meeting the global food demand of the future by engineering crop photosynthesis and yield potential’, Cell, 161(1), pp. 56–66. Available at: https://doi.org/10.1016/j.cell.2015.03.019.

Lorimer, G.H., Andrews, T.J. and Tolbert, N.E. (1973) ‘Ribulose Diphosphate Oxygenase. II. Further Proof of Reaction Products and Mechanism of Action’, Biochemistry, 12(1), pp. 18–23. Available at: https://doi.org/10.1021/bi00725a004.

Lorimer, G.H., Badger, M.R. and Andrews, T.J. (1976) ‘The Activation of Ribulose-1,5-bisphosphate Carboxylase by Carbon Dioxide and Magnesium Ions. Equilibria, Kinetics, a Suggested Mechanism, and Physiological implications’, Biochemistry, 15(3), pp. 529–536. Available at: https://doi.org/10.1021/bi00648a012.

Lorimer, G.H. and Miziorko, H.M. (1980) ‘Carbamate formation on the epsilon-amino group of a lysyl residue as the basis for the activation of ribulosebisphosphate carboxylase by CO2 and Mg2+.’, Biochemistry, 19(23), pp. 5321–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6778504.

Lowe, N.M. (2021) ‘The global challenge of hidden hunger: Perspectives from the field’, Proceedings of the Nutrition Society, 80(3), pp. 283–289. Available at: https://doi.org/10.1017/S0029665121000902.

Lüttge, U. (2004) ‘Ecophysiology of Crassulacean Acid Metabolism (CAM)’, Annals of Botany. Oxford University Press, pp. 629–652. Available at: https://doi.org/10.1093/aob/mch087.

Lyons, T.W., Reinhard, C.T. and Planavsky, N.J. (2014) ‘The rise of oxygen in Earth’s early ocean and atmosphere’, Nature, 506(7488), pp. 307–315. Available at: https://doi.org/10.1038/nature13068.

MacAlister, C.A., Ohashi-Ito, K. and Bergmann, D.C. (2007) ‘Transcription factor control of asymmetric cell divisions that establish the stomatal lineage’, Nature, 445(7127), pp. 537–540. Available at: https://doi.org/10.1038/nature05491.

MacRobbie, E.A.C. and Lettau, J. (1980) ‘Ion content and aperture in “isolated” guard cells of Commelina communis L.’, The Journal of Membrane Biology, 53(3), pp. 199–205. Available at: https://doi.org/10.1007/BF01868825.

Maia, L.B. et al. (2016) ‘Reduction of Carbon Dioxide by a Molybdenum-Containing Formate Dehydrogenase: A Kinetic and Mechanistic Study’, Journal of the American Chemical Society, 138(28), pp. 8834–8846. Available at: https://doi.org/10.1021/jacs.6b03941.

Malhi, G.S., Kaur, M. and Kaushik, P. (2021) ‘Impact of climate change on agriculture and its mitigation strategies: A review’, Sustainability (Switzerland), 13(3), pp. 1–21. Available at: https://doi.org/10.3390/su13031318.

Martin-Avila, E. et al. (2020) ‘Modifying plant photosynthesis and growth via simultaneous chloroplast transformation of rubisco large and small subunits’, Plant Cell, 32(9), pp. 2898–2916. Available at: https://doi.org/10.1105/TPC.20.00288.

Masle, J., Gilmore, S.R. and Farquhar, G.D. (2005) ‘The ERECTA gene regulates plant transpiration efficiency in Arabidopsis’, Nature, 436(7052), pp. 866–870. Available at: https://doi.org/10.1038/nature03835.

Matsumura, H. et al. (2020) ‘Hybrid Rubisco with Complete Replacement of Rice Rubisco Small Subunits by Sorghum Counterparts Confers C4 Plant-like High Catalytic Activity’, Molecular Plant, 13(11), pp. 1570–1581. Available at: https://doi.org/10.1016/j.molp.2020.08.012.

McAusland, L. et al. (2016) ‘Effects of kinetics of light-induced stomatal responses on photosynthesis and water-use efficiency’, The New phytologist, 211(4), pp. 1209–1220. Available at: https://doi.org/10.1111/nph.14000.

McCulloch, M.T. et al. (2024) ‘300 years of sclerosponge thermometry shows global warming has exceeded 1.5 °C’, Nature Climate Change, 14(2), pp. 171–177. Available at: https://doi.org/10.1038/s41558-023-01919-7.

McElwain, J.C. and Chaloner, W.G. (1995) ‘Stomatal density and index of fossil plants track atmospheric carbon dioxide in the palaeozoic’, Annals of Botany, 76(4), pp. 389–395. Available at: https://doi.org/10.1006/anbo.1995.1112.

McFadden, B.A. and Small, C.L. (1988) ‘Cloning, expression and directed mutagenesis of the genes for ribulose bisphosphate carboxylase/oxygenase’, Photosynthesis Research, 18(1–2), pp. 245–260. Available at: https://doi.org/10.1007/BF00042987.

Miglietta, F. et al. (2001) ‘Free-air CO2 enrichment (FACE) of a poplar plantation: The POPFACE fumigation system’, New Phytologist, 150(2), pp. 465–476. Available at: https://doi.org/10.1046/j.1469-8137.2001.00115.x.

Miller, F.P. (2008) ‘After 10,000 Years of Agriculture, Whither Agronomy?’, Agronomy Journal, 100(1), p. 22. Available at: https://doi.org/10.2134/agrojnl2007.0013.

Mohammed, U. et al. (2019) ‘Rice plants overexpressing OsEPF1 show reduced stomatal density and increased root cortical aerenchyma formation’, Scientific Reports, 9(1). Available at: https://doi.org/10.1038/s41598-019-41922-7.

Molisch, H. (1907) Die purpurbakterien : nach neuen untersuchungen; eine mikrobiologische studie, Die purpurbakterien : nach neuen untersuchungen; eine mikrobiologische studie /. Jena: Fischer. Available at: https://doi.org/10.5962/bhl.title.115708.

Monteith, J.L. (1977) ‘Climate and the efficiency of crop production in Britain’, Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 281(980), pp. 277–294. Available at: https://doi.org/10.1098/rstb.1977.0140.

Moore, B.D. et al. (1992) ‘Identification and levels of 2′-carboxyarabinitol in leaves’, Plant Physiology, 99(4), pp. 1546–1550. Available at: https://doi.org/10.1104/pp.99.4.1546.

Moore, B.D., Kobza, J. and Seemann, J.R. (1991) ‘Measurement of 2-carboxyarabinitol 1-phosphate in plant leaves by isotope dilution’, Plant Physiology, 96(1), pp. 208–213. Available at: https://doi.org/10.1104/pp.96.1.208.

Moore, B.D. and Seemann, J.R. (1992) ‘Metabolism of 2′-carboxyarabinitol in leaves’, Plant Physiology, 99(4), pp. 1551–1555. Available at: https://doi.org/10.1104/pp.99.4.1551.

Morita, K. et al. (2014) ‘Unusual small subunit that is not expressed in photosynthetic cells alters the catalytic properties of Rubisco in rice’, Plant Physiology, 164(1), pp. 69–79. Available at: https://doi.org/10.1104/pp.113.228015.

Mueller-Cajar, O. et al. (2011) ‘Structure and function of the AAA + protein CbbX, a red-type Rubisco activase’, Nature, 479(7372), pp. 194–199. Available at: https://doi.org/10.1038/nature10568.

Mueller-Cajar, O., Stotz, M. and Bracher, A. (2014) ‘Maintaining photosynthetic CO2 fixation via protein remodelling: The Rubisco activases’, Photosynthesis Research, 119(1–2), pp. 191–201. Available at: https://doi.org/10.1007/s11120-013-9819-0.

Mulkidjanian, A.Y. et al. (2006) 'The cyanobacterial genome core and the origin of photosynthesis', Proceedings of the National Academy of Sciences of the United States of America, 103(35), pp. 13126–13131. Available at: https://doi.org/10.1073/pnas.0605709103.

Murphy, D.J. (2007) People, Plants and Genes: The Story of Crops and Humanity, People, Plants and Genes: The Story of Crops and Humanity. Available at: https://doi.org/10.1093/acprof:oso/9780199207145.001.0001.

Nelson, N. and Ben-Shem, A. (2004) 'The complex architecture of oxygenic photosynthesis', Nature Reviews Molecular Cell Biology, 5(12), pp. 971–982. Available at: https://doi.org/10.1038/nrm1525.

Nelson, N. and Yocum, C.F. (2006) 'Structure and function of photosystems I and II', Annual Review of Plant Biology, 57, pp. 521–565. Available at: https://doi.org/10.1146/annurev.arplant.57.032905.105350.

Nguyen, T.B.A. et al. (2023) 'Engineering stomata for enhanced carbon capture and water-use efficiency', Trends in Plant Science, 28(11), pp. 1290–1309. Available at: https://doi.org/10.1016/j.tplants.2023.06.002.

Nilson, S. and Assmann, S. (2007) 'The control of transpiration. Insights from Arabidopsis', Plant Physiology, 143, pp. 19–27.

O'Brien, P.J. and Herschlag, D. (1999) 'Catalytic promiscuity and the evolution of new enzymatic activities', Chemistry and Biology, 6(4). Available at: https://doi.org/10.1016/S1074-5521(99)80033-7.

Orr, D.J. et al. (2023) 'Dynamics of Rubisco regulation by sugar phosphate derivatives and their phosphatases', Journal of Experimental Botany, 74(2), pp. 581–590. Available at: https://doi.org/10.1093/jxb/erac386.

Outlaw, W.H. (1983) 'Current concepts on the role of potassium in stomatal movements', Physiologia Plantarum, 59(2), pp. 302–311. Available at: https://doi.org/10.1111/j.1399-3054.1983.tb00775.x.

Papineau, D., Mojzsis, S.J. and Schmitt, A.K. (2007) 'Multiple sulfur isotopes from Paleoproterozoic Huronian interglacial sediments and the rise of atmospheric oxygen', Earth and Planetary Science Letters, 255(1–2), pp. 188–212. Available at: https://doi.org/10.1016/j.epsl.2006.12.015.

Parry, M.A.J. et al. (2008) 'Rubisco regulation: A role for inhibitors', Journal of Experimental Botany, 59(7), pp. 1569–1580. Available at: https://doi.org/10.1093/jxb/ern084.

Peterhansel, C. et al. (2010) 'Photorespiration', The Arabidopsis Book, 8(8), p. e0130. Available at: https://doi.org/10.1199/tab.0130.

Pillitteri, L.J. et al. (2007) 'Termination of asymmetric cell division and differentiation of stomata', Nature, 445(7127), pp. 501–505. Available at: https://doi.org/10.1038/nature05467.

Planavsky, N.J. et al. (2014) 'Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event', Nature Geoscience, 7(4), pp. 283–286. Available at: https://doi.org/10.1038/ngeo2122.

Plusquellec, H. (2002) 'Is the daunting challenge of irrigation achievable?', Irrigation and Drainage, 51(3), pp. 185–198. Available at: https://doi.org/10.1002/ird.51.

Poffenroth, M., Green, D.B. and Tallman, G. (1992) 'Sugar concentrations in guard cells of vicia faba illuminated with red or blue light: Analysis by high performance liquid chromatography', Plant Physiology, 98(4), pp. 1460–1471. Available at: https://doi.org/10.1104/pp.98.4.1460.

Price, G.D. et al. (1995) 'Specific reduction of chloroplast glyceraldehyde-3-phosphate dehydrogenase activity by antisense RNA reduces CO2 assimilation via a reduction in ribulose bisphosphate regeneration in transgenic tobacco plants', Planta, 195(3), pp. 369–378. Available at: https://doi.org/10.1007/BF00202594.

Prywes, N. et al. (2022) 'Rubisco function, evolution, and engineering'. Available at: http://arxiv.org/abs/2207.10773.

Qi, X. and Torii, K.U. (2018) 'Hormonal and environmental signals guiding stomatal development', BMC Biology, 16(1). Available at: https://doi.org/10.1186/s12915-018-0488-5.

Quarrie, S.A. and Jones, H.G. (1977) 'Effects of abscisic acid and water stress on development and morphology of wheat', Journal of Experimental Botany, 28(1), pp. 192–203. Available at: https://doi.org/10.1093/jxb/28.1.192.

Le Quéré, C. et al. (2016) 'Global carbon budget 2016', Earth System Science Data, 8, pp. 605–649.

Rasmussen, B., Bekker, A. and Fletcher, I.R. (2013) 'Correlation of Paleoproterozoic glaciations based on U-Pb zircon ages for tuff beds in the Transvaal and Huronian Supergroups', Earth and Planetary Science Letters, 382, pp. 173–180. Available at: https://doi.org/10.1016/j.epsl.2013.08.037.

Raven, J.A. (2002) 'Selection pressures on stomatal evolution', New Phytologist, 153(3), pp. 371–386. Available at: https://doi.org/10.1046/j.0028-646X.2001.00334.x.

Ray, D.K. et al. (2012) 'Recent patterns of crop yield growth and stagnation', Nature Communications, 3. Available at: https://doi.org/10.1038/ncomms2296.

Ray, D.K. et al. (2013) 'Yield Trends Are Insufficient to Double Global Crop Production by 2050', PLoS ONE, 8(6). Available at: https://doi.org/10.1371/journal.pone.0066428.

Raza, S. and Ameen, A. (2017) 'Green Revolution: A Review', International Journal of Advances in Scientific Research, 3(08), pp. 90–95. Available at: www.ssjournals.com.

Read, B.A. and Tabita, F.R. (1992a) 'A Hybrid Ribulosebisphosphate Carboxylase/Oxygenase Enzyme Exhibiting a Substantial Increase in Substrate Specificity Factor', Biochemistry, 31(24), pp. 5553–5560. Available at: https://doi.org/10.1021/bi00139a018.

Read, B.A. and Tabita, F.R. (1992b) 'Amino Acid Substitutions in the Small Subunit of Ribulose-1, 5-bisphosphate Carboxylase/Oxygenase That Influence Catalytic Activity of the Holoenzyme', Biochemistry, 31(2), pp. 519–525. Available at: https://doi.org/10.1021/bi00117a031.

Reckmann, U., Scheibe, R. and Raschke, K. (1990) 'Rubisco activity in guard cells compared with the solute requirement for stomatal opening', Plant Physiology, 92(1), pp. 246–253. Available at: https://doi.org/10.1104/pp.92.1.246.

Richards, R.A. (2000) 'Selectable traits to increase crop photosynthesis and yield of grain crops', Journal of Experimental Botany, 51(SPEC. ISS.), pp. 447–458. Available at: https://doi.org/10.1093/jexbot/51.suppl_1.447.

Ripple, W.J. et al. (2023) 'Many risky feedback loops amplify the need for climate action', One Earth, 6(2), pp. 86–91. Available at: https://doi.org/10.1016/j.oneear.2023.01.004.

Robinson, S.P. and Portis, A.R. (1988) 'Release of the nocturnal inhibitor, carboxyarabinitol-1-phosphate, from ribulose bisphosphate carboxylase/oxygenase by rubisco activase', FEBS Letters, 233(2), pp. 413–416. Available at: https://doi.org/10.1016/0014-5793(88)80473-3.

Sadekar, S., Raymond, J. and Blankenship, R.E. (2006) 'Conservation of distantly related membrane proteins: Photosynthetic reaction centers share a common structural core', Molecular Biology and Evolution, 23(11), pp. 2001–2007. Available at: https://doi.org/10.1093/molbev/msl079.

Sage, R.F., Sage, T.L. and Kocacinar, F. (2012) 'Photorespiration and the evolution of C4 photosynthesis', Annual Review of Plant Biology, pp. 19–47. Available at: https://doi.org/10.1146/annurev-arplant-042811-105511.

Sakoda, K. et al. (2020) 'Higher Stomatal Density Improves Photosynthetic Induction and Biomass Production in Arabidopsis Under Fluctuating Light', Frontiers in Plant Science, 11. Available at: https://doi.org/10.3389/fpls.2020.589603.

Sakoda, K. et al. (2021) 'Effects of introduction of sorghum RbcS with rice RbcS knockdown by RNAi on photosynthetic activity and dry weight in rice', Plant Production Science, 24(3), pp. 346–353. Available at: https://doi.org/10.1080/1343943X.2020.1847669.

Saschenbrecker, S. et al. (2007) 'Structure and Function of RbcX, an Assembly Chaperone for Hexadecameric Rubisco', Cell, 129(6), pp. 1189–1200. Available at: https://doi.org/10.1016/j.cell.2007.04.025.

Sato, T., Atomi, H. and Imanaka, T. (2007) 'Archaeal type III RuBisCOs function in a pathway for AMP metabolism', Science, 315(5814), pp. 1003–1006. Available at: https://doi.org/10.1126/science.1135999.

Saysel, A.K. and Barlas, Y. (2001) 'A dynamic model of salinization on irrigated lands', Ecological Modelling, 139(2–3), pp. 177–199. Available at: https://doi.org/10.1016/S0304-3800(01)00242-3.

Schidlowski, M. (1988) 'A 3,800-million-year isotopic record of life from carbon in sedimentary rocks', Nature, 333(6171), pp. 313–318. Available at: https://doi.org/10.1038/333313a0.

Schlüter, U. et al. (2003) 'Photosynthetic performance of an Arabidopsis mutant with elevated stomatal density (sdd1-1) under different light regimes', Journal of Experimental Botany, 54(383), pp. 867–874. Available at: https://doi.org/10.1093/jxb/erg087.

Schluter, U. and Weber, A.P.M. (2020) 'Regulation and Evolution of C4 Photosynthesis', Annual Review of Plant Biology, 71, pp. 183–215. Available at: https://doi.org/10.1146/annurev-arplant-042916-040915.

Schnabl, H. and Raschke, K. (1980) ' Potassium Chloride as Stomatal Osmoticum in Allium cepa L., a Species Devoid of Starch in Guard Cells ', Plant Physiology, 65(1), pp. 88–93. Available at: https://doi.org/10.1104/pp.65.1.88.

Schneider, G., Lindqvist, Y. and Brändén, C.I. (1992) 'RUBISCO: Structure and mechanism', Annual Review of Biophysics and Biomolecular Structure, pp. 119–143. Available at: https://doi.org/10.1146/annurev.bb.21.060192.001003.

Schopf, J.W. (2011) 'The paleobiological record of photosynthesis', Photosynthesis Research, 107(1), pp. 87–101. Available at: https://doi.org/10.1007/s11120-010-9577-1.

Schroeder, J.I. et al. (2001) 'Guard cell signal transduction', Annual Review of Plant Biology, 52, pp. 627–658. Available at: https://doi.org/10.1146/annurev.arplant.52.1.627.

Schuchmann, K. and Müller, V. (2013) 'Direct and reversible hydrogenation of CO2to formate by a bacterial carbon dioxide reductase', Science, 342(6164), pp. 1382–1385. Available at: https://doi.org/10.1126/science.1244758.

Schuler, M.L. et al. (2018) 'SHORTROOT-mediated increase in stomatal density has no impact on photosynthetic efficiency', Plant Physiology, 176(1), pp. 757–772. Available at: https://doi.org/10.1104/pp.17.01005.

Schulz, L. et al. (2022) 'Evolution of increased complexity and specificity at the dawn of form I Rubiscos', Science, 378(6616). Available at: https://doi.org/10.1126/science.abq1416.

Schulz, L., Sendker, F.L. and Hochberg, G.K.A. (2022) 'Non-adaptive complexity and biochemical function', Current Opinion in Structural Biology, 73. Available at: https://doi.org/10.1016/j.sbi.2022.102339.

Schwartz, R.M. and Dayhoff, M.O. (1978) 'Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts', Science, 199(4327), pp. 395–403. Available at: https://doi.org/10.1126/science.202030.

Sekine, Y. et al. (2011) 'Manganese enrichment in the Gowganda Formation of the Huronian Supergroup: A highly oxidizing shallow-marine environment after the last Huronian glaciation', Earth and Planetary Science Letters, 307(1–2), pp. 201–210. Available at: https://doi.org/10.1016/j.epsl.2011.05.001.

Sharkey, T.D. (1988) 'Estimating the rate of photorespiration in leaves', Physiologia Plantarum, 73(1), pp. 147–152. Available at: https://doi.org/10.1111/j.1399-3054.1988.tb09205.x.

Sharkey, T.D. (2023) 'The discovery of rubisco', Journal of Experimental Botany, 74(2), pp. 510–519. Available at: https://doi.org/10.1093/jxb/erac254.

Sharkey, T.D., Badger, M.R. and Andrews, T.J. (2001) 'Increased heat sensitivity of photosynthesis in tobacco plants with reduced Rubisco activase', pp. 147–156.

Sharwood, R.E. (2017) 'Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops', New Phytologist, pp. 494–510. Available at: https://doi.org/10.1111/nph.14351.

Shaw, G.H. (2016) 'The Origin of Oxygenic Photosynthesis and Its Impact on the Atmosphere', in Earth's Early Atmosphere and Oceans, and The Origin of Life, pp. 75–93. Available at: https://doi.org/10.1007/978-3-319-21972-1_11.

Siebert, S. et al. (2010) 'Groundwater use for irrigation - A global inventory', Hydrology and Earth System Sciences, 14(10), pp. 1863–1880. Available at: https://doi.org/10.5194/hess-14-1863-2010.

Simkin, A.J., López-Calcagno, P.E. and Raines, C.A. (2019) 'Feeding the world: improving photosynthetic efficiency for sustainable crop production', Journal of Experimental Botany, 70(4), pp. 1119–1140. Available at: https://doi.org/10.1093/JXB/ERY445.

Smil, V. (2002) 'Nitrogen and Food Production: Proteins for Human Diets', AMBIO: A Journal of the Human Environment, 31(2), p. 126. Available at: https://doi.org/10.1639/0044-7447(2002)031[0126:nafppf]2.0.co;2.

Spreitzer, R.J. and Salvucci, M.E. (2002) 'Rubisco: Structure, regulatory interactions, and possibilities for a better enzyme', Annual Review of Plant Biology, 53, pp. 449–475. Available at: https://doi.org/10.1146/annurev.arplant.53.100301.135233.

Stec, B. (2012) 'Structural mechanism of RuBisCO activation by carbamylation of the active site lysine', Proceedings of the National Academy of Sciences of the United States of America, 109(46), pp. 18785–18790. Available at: https://doi.org/10.1073/pnas.1210754109.

Still, C.J. et al. (2003) 'Global distribution of C3 and C4 vegetation: Carbon cycle implications', Global Biogeochemical Cycles, pp. 6-1-6–14. Available at: https://doi.org/10.1029/2001gb001807.

Stocker, T.F. et al. (2013) 'Climate change 2013 the physical science basis: Working Group I contribution to the fifth assessment report of the intergovernmental panel on climate change', Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 9781107057, pp. 1–1535. Available at: https://doi.org/10.1017/CBO9781107415324.

Szabo, S. (2016) 'Urbanisation and Food Insecurity Risks: Assessing the Role of Human Development', Oxford Development Studies, 44(1), pp. 28–48. Available at: https://doi.org/10.1080/13600818.2015.1067292.

Tabita, F.R. et al. (2008) 'Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships', Journal of Experimental Botany, 59(7), pp. 1515–1524. Available at: https://doi.org/10.1093/jxb/erm361.

Talbott, L.D. and Zeiger, E. (1993) 'Sugar and organic acid accumulation in guard cells of Vicia faba in response to red and blue light', Plant Physiology, 102(4), pp. 1163–1169. Available at: https://doi.org/10.1104/pp.102.4.1163.

Talbott, L.D. and Zeiger, E. (1996) 'Central roles for potassium and sucrose in guard-cell osmoregulation', Plant Physiology, 111(4), pp. 1051–1057. Available at: https://doi.org/10.1104/pp.111.4.1051.

Tallman, G. and Zeiger, E. (1988) 'Light quality and osmoregulation in vicia guard cells : evidence for involvement of three metabolic pathways.', Plant physiology, 88(3), pp. 887–95. Available at: https://doi.org/10.1104/pp.88.3.887.

Tanaka, Y. et al. (2013) 'Enhancement of leaf photosynthetic capacity through increased stomatal density in Arabidopsis', New Phytologist, 198(3), pp. 757–764. Available at: https://doi.org/10.1111/nph.12186.

Tenenbaum, D.J. (2008) 'Food vs. fuel diversion of crops could cause more hunger', Environmental Health Perspectives, 116(6). Available at: https://doi.org/10.1289/ehp.116-a254.

The United Nations (2015) 'The Paris Agreement', pp. 1–16. Available at: https://unfccc.int/sites/default/files/english_paris_agreement.pdf.

Tice, M.M. and Lowe, D.R. (2004) 'Photosynthetic microbial mats in the 3,416-Myr-old ocean', Nature, 431(7008), pp. 549–552. Available at: https://doi.org/10.1038/nature02888.

Tice, M.M. and Lowe, D.R. (2006) 'Hydrogen-based carbon fixation in the earliest known photosynthetic organisms', Geology, 34(1), pp. 37–40. Available at: https://doi.org/10.1130/G22012.1.

Tsai, Y.C.C. et al. (2012) 'Chaperonin cofactors, Cpn10 and Cpn20, of green algae and plants function as hetero-oligomeric ring complexes', Journal of Biological Chemistry, 287(24), pp. 20471–20481. Available at: https://doi.org/10.1074/jbc.M112.365411.

Tsai, Y.C.C. et al. (2015) 'Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria', Nature Communications, 6. Available at: https://doi.org/10.1038/ncomms9883.

van der Vies, S.M., Bradley, D. and Gatenby, A.A. (1986) ' Assembly of cyanobacterial and higher plant ribulose bisphosphate carboxylase subunits into functional homologous and heterologous enzyme molecules in Escherichia coli ', The EMBO Journal, 5(10), pp. 2439–2444. Available at: https://doi.org/10.1002/j.1460-2075.1986.tb04519.x.

Vigna, L. et al. (1980) 'Dependence of the Stomatal Index on Environmental Factors during Stomatal Differentiation in Leaves of Vigna sinensis L.', Journal of Experimental Botany, 31(124), pp. 1211–1216.

Voordouw, G. et al. (1987) 'Site-directed mutagenesis of the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from Anacystis nidulans', European Journal of Biochemistry, 163(3), pp. 591–598. Available at: https://doi.org/10.1111/j.1432-1033.1987.tb10908.x.

Walker, B.J. et al. (2016) 'The Costs of Photorespiration to Food Production Now and in the Future', Annual Review of Plant Biology, 67, pp. 107–129. Available at: https://doi.org/10.1146/annurev-arplant-043015-111709.

Walker, J.C.G. et al. (1983) 'Environmental evolution of the Archaean-Early Proterozoic Earth', in Earth's Earliest Biosphere: Its Origin and Evolution, pp. 260–290.

Wang, C. et al. (2016) 'PdEPF1 regulates water-use efficiency and drought tolerance by modulating stomatal density in poplar', Plant Biotechnology Journal, 14(3), pp. 849–860. Available at: https://doi.org/10.1111/pbi.12434.

Wang, Y.L. et al. (2001) 'Properties of hybrid enzymes between Synechococcus large subunits and higher plant small subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase in Escherichia coli', Archives of Biochemistry and Biophysics, 396(1), pp. 35–42. Available at: https://doi.org/10.1006/abbi.2001.2555.

Ward, L.M., Rasmussen, B. and Fischer, W.W. (2019) 'Primary Productivity Was Limited by Electron Donors Prior to the Advent of Oxygenic Photosynthesis', Journal of Geophysical Research: Biogeosciences, 124(2), pp. 211–226. Available at: https://doi.org/10.1029/2018JG004679.

Watson, G.M.F., Yu, J.P. and Tabita, F.R. (1999) 'Unusual ribulose 1,5-bisphosphate carboxylase/oxygenase of anoxic Archaea', Journal of Bacteriology, 181(5), pp. 1569–1575. Available at: https://doi.org/10.1128/jb.181.5.1569-1575.1999.

Weigel, H.J. and Manderscheid, R. (2012) 'Crop growth responses to free air CO 2 enrichment and nitrogen fertilization: Rotating barley, ryegrass, sugar beet and wheat', European Journal of Agronomy, 43, pp. 97–107. Available at: https://doi.org/10.1016/j.eja.2012.05.011.

Wheatley, N.M. et al. (2014) 'Structure and identification of a Pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (RuBisCO) assembly factor in the α-carboxysome', Journal of Biological Chemistry, 289(11), pp. 7973–7981. Available at: https://doi.org/10.1074/jbc.M113.531236.

Williford, K.H. et al. (2011) 'Constraining atmospheric oxygen and seawater sulfate concentrations during Paleoproterozoic glaciation: In situ sulfur three-isotope microanalysis of pyrite from the Turee Creek Group, Western Australia', Geochimica et Cosmochimica Acta, 75(19), pp. 5686–5705. Available at: https://doi.org/10.1016/j.gca.2011.07.010.

Wilson, R.H. et al. (2018) 'An improved Escherichia coli screen for Rubisco identifies a protein-protein interface that can enhance CO2-fixation kinetics', Journal of Biological Chemistry, 293(1), pp. 18–27. Available at: https://doi.org/10.1074/jbc.M117.810861.

Wilson, R.H. and Hayer-Hartl, M. (2018) 'Complex Chaperone Dependence of Rubisco Biogenesis', Biochemistry, 57(23), pp. 3210–3216. Available at: https://doi.org/10.1021/acs.biochem.8b00132.

Wingler, A. et al. (2000) 'Photorespiration: Metabolic pathways and their role in stress protection', Philosophical Transactions of the Royal Society B: Biological Sciences, 355(1402), pp. 1517–1529. Available at: https://doi.org/10.1098/rstb.2000.0712.

Witt, H.T., Mueller, A. and Rumberg, B. (1961) 'Oxidized cytochrome and chlorophyll C2 + in photosynthesis', Nature, 192(4806), pp. 967–969. Available at: https://doi.org/10.1038/192967a0.

Woodward, F.I. (1987) 'Stomatal numbers are sensitive to increases in CO2 from pre-industrial levels', Nature, 327, pp. 617–618.

Woodward, F.I. and Bazzaz, F.A. (1988) 'The responses of stomatal density to CO2 partial pressure', Journal of Experimental Botany, 39(12), pp. 1771–1781. Available at: https://doi.org/10.1093/jxb/39.12.1771.

Woodward, F.I., Lake, J.A. and Quick, W.P. (2002) 'Stomatal development and CO2: Ecological consequences', New Phytologist, 153(3), pp. 477–484. Available at: https://doi.org/10.1046/j.0028-646X.2001.00338.x.

Wrighton, K.C. et al. (2016) 'RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria', ISME Journal, 10(11), pp. 2702–2714. Available at: https://doi.org/10.1038/ismej.2016.53.

Xiong, D. and Flexas, J. (2020) 'From one side to two sides: the effects of stomatal distribution on photosynthesis', New Phytologist, 228(6), pp. 1754–1766. Available at: https://doi.org/10.1111/nph.16801.

Yabe, S. et al. (2022) ' Vulcanimicrobium alpinus gen. nov. sp. nov., the first cultivated representative of the candidate phylum "Eremiobacterota", is a metabolically versatile aerobic anoxygenic phototroph ', ISME Communications, 2(1). Available at: https://doi.org/10.1038/s43705-022-00201-9.

Yan, X. et al. (2018) 'GroEL Ring Separation and Exchange in the Chaperonin Reaction', Cell, 172(3), pp. 605-617.e11. Available at: https://doi.org/10.1016/j.cell.2017.12.010.

Yang, M. and Sack, F.D. (1995) 'The too many mouths and four lips mutations affect stomatal production in arabidopsis', Plant Cell, 7(12), pp. 2227–2239. Available at: https://doi.org/10.2307/3870164.

Yang, X. et al. (2024) 'Engineering Crassulacean Acid Metabolism in C3 and C4 Plants', Cold Spring Harbor Perspectives in Biology, 16(4). Available at: https://doi.org/10.1101/cshperspect.a041674.

Yeats, T.H. and Rose, J.K.C. (2013) 'The Formation and Function of Plant Cuticles', Plant physiology, 163, pp. 5–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10444101.

Yin, X. et al. (2017) 'CRISPR-Cas9 and CRISPR-Cpf1 mediated targeting of a stomatal developmental gene EPFL9 in rice', Plant Cell Reports, 36(5), pp. 745–757. Available at: https://doi.org/10.1007/s00299-017-2118-z.

Yoo, C.Y. et al. (2010) 'The Arabidopsis GTL1 transcription factor regulates water use efficiency and drought tolerance by modulating stomatal density via transrepression of SDD1', Plant Cell, 22(12), pp. 4128–4141. Available at: https://doi.org/10.1105/tpc.110.078691.

Zhao, J. et al. (2024) 'Genome-wide analysis of the TIFY family in Lycium and the negative regulation of stomatal development by LrJAZ2 gene', Plant Physiology and Biochemistry, 206. Available at: https://doi.org/10.1016/j.plaphy.2023.108285.

Zhao, M. et al. (2023) 'Oxygenation of the Earth aided by mineral–organic carbon preservation', Nature Geoscience, 16(3), pp. 262–267. Available at: https://doi.org/10.1038/s41561-023-01133-2.

Zhao, Q. and Liu, C. (2018) 'Chloroplast chaperonin: An intricate protein folding machine for photosynthesis', Frontiers in Molecular Biosciences, 4(JAN). Available at: https://doi.org/10.3389/fmolb.2017.00098.

Zhu, X.G., Long, S.P. and Ort, D.R. (2008) ‘What is the maximum efficiency with which photosynthesis can convert solar energy into biomass?’, Current Opinion in Biotechnology, pp. 153–159. Available at: https://doi.org/10.1016/j.copbio.2008.02.004.

Zhu, X.G., Long, S.P. and Ort, D.R. (2010) ‘Improving photosynthetic efficiency for greater yield’, Annual Review of Plant Biology, 61, pp. 235–261. Available at: https://doi.org/10.1146/annurev-arplant-042809-112206.

Chapter 2. Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off

Chapter 2, Contributors:

Bouvier, J.W., Emms, D.M., Rhodes, T., Bolton, J.S., Brasnett, A., Eddershaw, A., Nielsen, J.R., Unitt, A., Whitney, S.M. and Kelly, S.

Author contributions to data Chapter 2:

S.K. conceived the study. J.W.B. conducted the analysis. J.W.B., T.R., S.M.W., D.M.E., and S.K. analyzed the data. J.W.B. and S.K. wrote the manuscript. J.R.N., J.S.B., A.E., A.B., and A.U. contributed to discussions and methodological development. All authors edited the manuscript.

This thesis chapter has been published at the following location:

Bouvier, J.W., et al., 2021. Molecular Biology and Evolution, 38(7), 2880-2896

All supplemental data files are provided in the online version of this manuscript.

2.1 Abstract

Rubisco assimilates CO_2 to form the sugars that fuel life on earth. Correlations between rubisco kinetic traits across species have led to the proposition that rubisco adaptation is highly constrained by catalytic trade-offs. However, these analyses did not consider the phylogenetic context of the enzymes that were analysed. Thus, it is possible that the correlations observed were an artefact of the presence of phylogenetic signal in rubisco kinetics and the phylogenetic relationship between the species that were sampled. Here, we conducted a phylogenetically-resolved analysis of rubisco kinetics and show that there is a significant phylogenetic signal in rubisco kinetic traits. We re-

evaluated the extent of catalytic trade-offs accounting for this phylogenetic signal and found that all were attenuated. Following phylogenetic correction, the largest catalytic trade-offs were observed between the Michaelis constant for CO_2 and carboxylase turnover (~21-37 %), and between the Michaelis constants for CO_2 and O_2 (~9-19 %), respectively. All other catalytic trade-offs were substantially attenuated such that they were marginal (<9 %) or non-significant. This phylogenetically resolved analysis of rubisco kinetic evolution also identified kinetic changes that occur concomitant with the evolution of C_4 photosynthesis. Finally, we show that phylogenetic constraints have played a larger role than catalytic trade-offs in limiting the evolution of rubisco kinetics. Thus, although there is strong evidence for some catalytic trade-offs, rubisco adaptation has been more limited by phylogenetic constraint than by the combined action of all such trade-offs.

2.2 Introduction

The vast majority of organic carbon on Earth entered the biosphere via the catalytic pocket of rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase) (Beer *et al.*, 2010). Whilst there are several metabolic contexts in which this rubisco-mediated reaction can occur, the most important of these in terms of global net primary production is photosynthesis (Andersson and Backlund, 2008). Here, rubisco catalyses the initial step of the Calvin-Benson-Bassham reductive pentose phosphate pathway, catalysing the fixation of CO_2 onto the acceptor molecule ribulose 1,5-bisphosphate (RuBP) to ultimately synthesise sugars. There is a diverse array of rubisco forms found across the tree of life. Plants and some bacteria contain Form I rubisco which is composed of both large (RbcL) and small (RbcS) subunits (Schneider, Lindqvist and Brändén, 1992; Tabita *et al.*, 2008), while Form II, III and IV rubisco found in other lineages are comprised of just the large subunit (Tabita *et al.*, 2008; Banda *et al.*, 2020). Although Form I rubisco contains two subunits, only the large subunit is essential for catalysis (Andrews, 1988; Lee and Tabita, 1990; Whitney, Houtz and Alonso, 2011), while the small subunit has an indirect effect on catalytic properties and activity (Andrews, 1988; Lee and Tabita, 1990; Lee, Berka and Tabita, 1991; Read and Tabita, 1992b, 1992a; Spreitzer, Peddi and Satagopan, 2005; Ishikawa *et al.*, 2011; Joshi *et al.*, 2015; Fukayama *et al.*, 2019; Martin-Avila *et al.*, 2020).

Given that rubisco is the entry point for carbon into the global food chain it is perhaps unsurprising that it is the most abundant enzyme on Earth (Ellis, 1979) with a global mass of ~0.7 gigatons (Bar-On and Milo, 2019). However, this abundance is in part due to the inefficiency of rubisco as a catalyst. Specifically, rubisco has a low rate of CO_2 assimilation (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010) and is poorly able to discriminate CO_2 and O_2 (Ogren and Bowes, 1971) causing it to catalyse both a carboxylation and an oxygenation reaction (Bowes, Ogren and Hageman, 1971; Chollet, 1977; Sharkey, 2020). Rubisco-mediated oxygenation of RuBP results in the production of 2-phosphoglycolate, which must then be metabolised to recover carbon and avoid depletion of metabolite pools (Eckardt, 2005; Sharwood, 2017). In plants this carbon scavenging process is known as photorespiration, and consumes ATP and reducing power and also liberates ammonia which must be re-assimilated (Peterhansel *et al.*, 2010). Although the oxygenation reaction catalysed by rubisco is not thought to be deleterious in the anoxic environment prevalent when the enzyme first evolved (Nisbet *et al.*, 2007; Erb and Zarzycki, 2018), under current atmospheric conditions it can comprise a quarter of all rubisco reactions in terrestrial plants (Ehleringer *et al.*, 1991). Thus, despite oxygenation serving a number of beneficial functions (Busch, 2020), at its current rate it represents a substantial metabolic burden reducing the productivity of some plants by up to 50 % (Ogren, 1984; Bauwe *et al.*, 2012).

Given the high energic cost incurred by the rubisco oxygenation reaction, a number of photoautotrophic organisms have evolved mechanisms to reduce the frequency of its occurrence (Flamholz and Shih, 2020). Collectively referred to as CO_2-concentrating mechanisms, these function to increase the concentration of CO_2 relative to O_2 in the vicinity of rubisco and thus increase the relative frequency of carboxylation reactions (Meyer and Griffiths, 2013; Flamholz and Shih, 2020). These CO_2-concentrating mechanisms range in complexity from the carboxysome microcompartments in cyanobacteria (Kaplan *et al.*, 1991; Espie and Kimber, 2011), to the physical separation of primary CO_2 assimilation from rubisco-mediated photosynthetic CO_2 reduction in plants that conduct C_4 photosynthesis (Hatch, 1987; Sage, Sage and Kocacinar, 2012; Edwards, 2019). The observation that evolution has resulted in an array of CO_2-concentrating mechanisms, rather than improve the CO_2 specificity of rubisco, has led many to question whether altering the kinetics

of the enzyme is possible (Ogren, 1984; Parry *et al.*, 2007, 2013; Whitney, Houtz and Alonso, 2011; Sharwood, Ghannoum and Whitney, 2016; Araújo *et al.*, 2017; Sharwood, 2017; Sharkey, 2020). This proposition that rubisco specificity cannot be improved was supported by observations that the oxygenase and carboxylase activities of rubisco appear to be tightly linked (Badger and Lorimer, 1976; Chollet and Anderson, 1976). Subsequently, multiple studies have supported this suggestion by reporting strong antagonistic relationships between rubisco specificity ($S_{C/O}$), carboxylase turnover (k_{catC}) and the Michaelis constant (i.e., an inverse measure of substrate affinity for an enzyme) for CO_2 (K_C), as well as between K_C and the Michaelis constant for O_2 (K_O) (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010). Collectively, these studies have led to the hypothesis that severe kinetic trait trade-offs hamstring the inherent efficiency by which the enzyme can catalyse CO_2 fixation, and that contemporary rubisco are near perfectly adapted within this heavily constrained catalytic landscape (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010). However, new evidence has begun to question this paradigm of rubisco evolution. First, recent analyses of the correlative nature of rubisco kinetics has demonstrated that associations between kinetic traits are weakened when a large number of species are considered (Flamholz *et al.*, 2019; Iñiguez *et al.*, 2020). Furthermore, engineering efforts to alter rubisco kinetics have produced enzyme variants that deviate from proposed catalytic trade-offs between $S_{C/O}$, k_{catC} and K_C (Wilson *et al.*, 2018; Zhou and Whitney, 2019). Finally, an updated examination of rubisco kinetics in the context of other enzymes has shown that it is not as inefficient a catalyst as often assumed (Bathellier *et al.*, 2018). Thus, together these results indicate that rubisco kinetic traits are perhaps not as inextricably linked as originally thought, and suggest that there is scope for increasing the catalytic efficiency of the enzyme as has happened in nature for rubisco in some red algae (Andersson and Backlund, 2008; Gunn *et al.*, 2020).

Although the kinetic traits of rubisco appear to be correlated, there are flaws to inferring causality from this correlation. This is because previous analyses that have inferred correlations have generally assumed that measurements of rubisco kinetic traits in different species are independent (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010; Flamholz *et al.*, 2019; Iñiguez *et al.*, 2020). However, this assumption has rarely been formally tested (though see e.g., Orr *et al.,* 2016),

and is unlikely to be true because rubisco in all species are related to each other by descent from a single ancestral gene. This means that the gene sequences encoding the enzyme in closely related species are more similar than the gene sequences from species that are more distantly related, a feature of rubisco which has long been exploited in systematics and evolutionary analyses to serve as an accurate proxy for the phylogenetic relationship between species (Gielly and Taberlet, 1994; APG, 1998, 2016). As sequence variation determines kinetic variation, closely related enzymes would be expected to also have similar kinetics, with the extent of this similarity being dependent on the underlying tree describing the relationship between species. This phenomenon, which is known as phylogenetic signal, can cause spurious correlations in measured trait values between species unless the structure of the phylogenetic tree is taken into consideration (Felsenstein, 1985; Grafen, 1989; Pagel and Harvey, 1989; Garland, 2001). Thus, as previous analyses of rubisco kinetics have not assessed whether a phylogenetic signal exists in rubisco kinetic traits, nor accounted for any phylogenetic signal which may exist, it is possible that the observed catalytic trade-offs inferred from the presence of correlations are, either wholly or in part, an artefact caused by this phylogenetic signal.

Here, we assess the presence of a phylogenetic signal in rubisco kinetic traits to evaluate the extent to which rubisco kinetic evolution is constrained by both phylogenetic effects and catalytic trade-offs. We demonstrate that there is a significant phylogenetic signal in all rubisco kinetic traits. This means that the similarity of kinetic measurements between species varies as a function of their evolutionary distance, and thus kinetic measurements in different species are non-independent. When this phylogenetic signal is correctly accounted for by using phylogenetic least squares regression, we reveal that inferred catalytic trade-offs are weakened and that rubisco kinetic traits have been evolving largely independently of each other. Moreover, we find that phylogenetic constraints, most likely resulting from a slow rate of molecular evolution, have constrained rubisco kinetic evolution to a greater extent than catalytic trade-offs. This new insight offers encouragement to efforts which aim to increase yields in food, fibre and fuel crops by engineering rubisco variants with increased catalytic efficiency.

2.3 Results

2.3.1 Rubisco kinetic data

A dataset comprising kinetic measurements for rubisco isolated from different photoautotrophs was obtained from (Flamholz *et al.*, 2019). Measurements of specificity ($S_{C/O}$) for CO_2 relative to O_2 (i.e., the overall carboxylation/oxygenation ratio of rubisco under defined concentrations of CO_2 and O_2 gases) in this dataset were normalised in order to overcome discrepancies between values determined using an oxygen electrode assay (Parry, Keys and Gutteridge, 1989) and high precision gas-phase-controlled 3H-RuBP-fixation assays (Kane et al., 1994) (see methods). To begin, the interrogation of this data was focussed on the angiosperms because this was the group with the largest and most complete set of kinetic measurements, and to minimize any impact of long-branch effects (Su and Townsend, 2015). It was also restricted to those species with measurements of $S_{C/O}$, maximum carboxylase turnover rate per active site (k_{catC}), and the Michaelis constant (i.e., the substrate concentration at half-saturated catalysed rate) for both CO_2 (K_C) and O_2 (K_O). The Michaelis constant for CO_2 in 20.95 % O_2 air (K_C^{air}) was also inferred as a function of both K_C and K_O (see methods). Of the 137 angiosperms that satisfied these filtration criteria, 19 also had measurements of the Michaelis constant for RuBP (K_{RuBP}). From here on, these constants and rates are collectively termed kinetic traits, where $S_{C/O}$, k_{catC}, K_C and K_C^{air} are referred to as carboxylase-related kinetic traits, and K_O as the oxygenase-related kinetic trait.

2.3.2 A significant phylogenetic signal exists in rubisco carboxylase-related kinetic traits in angiosperms

Consistent with previous analyses (Flamholz *et al.*, 2019), all kinetic traits were log transformed to ensure they conformed to the distribution assumptions of the statistical analyses herein. To assess whether rubisco in different angiosperms display similar kinetics as a consequence of their phylogenetic relationship, the kinetic traits were analyzed in the context of the phylogenetic tree by which the species are related (Figure 2.1). Here, all kinetic traits were subject to interrogation for a phylogenetic signal (table 2.1) except for K_{RuBP}, which was omitted owing to the limited number of measurements available for this trait. For these analyses, several statistical tools varying in their

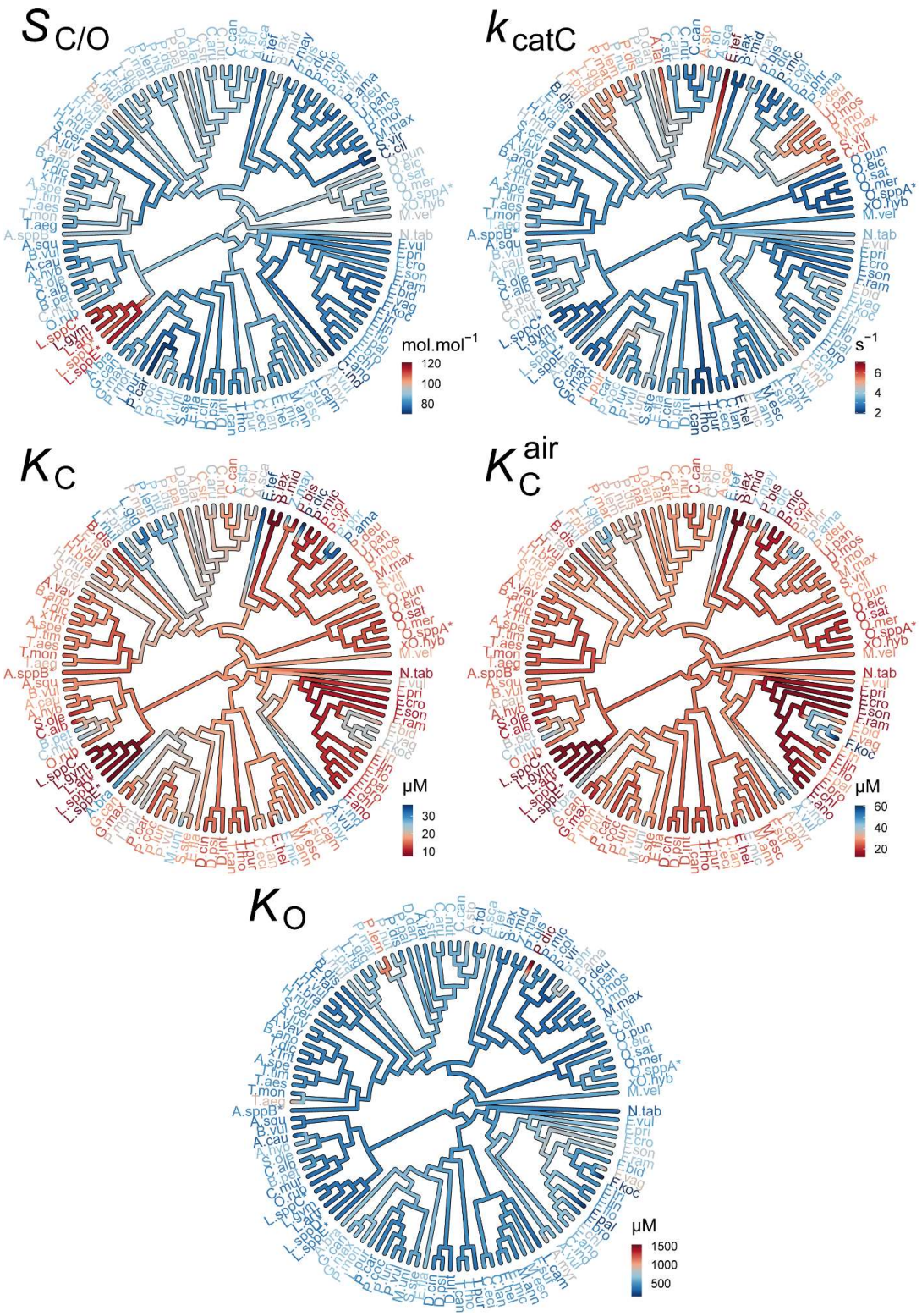


Figure 2.1. The evolution of rubisco kinetic traits in angiosperms. Phylogenetic tree of angiosperm species depicting the kinetic trait values in of the species used in this dataset and the maximum likelihood inferred ancestral kinetic traits for internal branches on the tree. Scale bars for colour schemes are presented next to each tree. Species names have been abbreviated for legibility and are provided in full in Appendix B [and ***Supplemental File 1 online***], table S4. $S_{C/O}$: specificity. k_{catC}: carboxylase turnover per site. K_C: the Michaelis constant for CO_2. K_C^{air} the inferred Michaelis constant for CO_2 in 20.95 % O_2 air. K_O: the Michaelis constant for O_2.

Table 2.1. The phylogenetic signal strength and associated significance level in rubisco kinetic traits in angiosperms using five different signal detection methods. Statistics are rounded to 3 decimal places and significance values are represented as α levels, where; α = 0.001 if $p < 0.001$, α = 0.01 if $0.001 < p < 0.01$, α = 0.05 if $0.01 < p < 0.05$, and α = ns if $p > 0.05$.

Kinetic Trait	C mean		I		K		K *		Lambda	
	Stat	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***
$S_{C/O}$	0.516	0.001	0.425	0.001	0.001	ns	0.001	ns	0.879	0.001
k_{catC}	0.350	0.001	0.224	0.001	0.003	0.001	0.003	0.001	0.968	0.001
K_C	0.282	0.001	0.248	0.001	0.001	0.01	0.002	0.01	0.902	0.001
K_C^{air}	0.234	0.001	0.242	0.001	0.001	0.05	0.001	0.05	0.392	ns
K_O	0.032	ns	-0.070	ns	0	ns	0	ns	0	ns

approach to phylogenetic signal detection were implemented and the presence or absence of a phylogenetic signal in each trait was judged by the majority result (i.e., the result of ≥ 3 out of 5 methods tested). Out of the methods utilized, Pagel's lambda (Pagel, 1999) and Blomberg's K and K* (Blomberg, Garland and Ives, 2003) analyse the distribution of trait values in extant species using an explicit Brownian motion model of trait evolution in which the traits evolve stochastically on the underlying phylogenetic tree at a uniform rate and independently among branches. In contrast, Moran's I (Gittleman and Kot, 1990) and Abouheif's Cmean (Abouheif, 1999) do not invoke any specific aspect of evolutionary theory, but instead test for a phylogenetic signal by assessing the correlation of trait values across evolutionary distance on the species tree using the concept of autocorrelation adopted from the field of spatial statistics (Cheverud, Dow and Leutenegger, 1985, 1986). For further discussion of the differences between these phylogenetic signal detection methods see (Münkemüller *et al.*, 2012).

Irrespective of the methodological approach used for inference, a significant phylogenetic signal was observed in all carboxylase-related kinetic traits of $S_{C/O}$, k_{catC}, K_C, and K_C^{air} in angiosperms (table 2.1; Figure 2.1). However, the strength of this signal varied across the different methods (table 2.1). In contrast, a phylogenetic signal was not detected in angiosperms for the oxygenase-related kinetic trait K_O (table 2.1; Figure 2.1). These measurements of phylogenetic signal were demonstrated to

not suffer from overfitting due to the use of the *rbcL* gene to infer the phylogenetic tree (Appendix B [and ***Supplemental File 1 online***], Figure S1; Figure S2; and table S1). Overall, this means that the similarity in carboxylase-related (but not oxygenase-related) kinetic traits in different angiosperms is dependent on their phylogenetic relationship. Therefore, conventional approaches to measure correlations that assume independence between observations of carboxylase-related kinetic trait values are incorrect, and correlation coefficients computed using such approaches have likely been over-estimated (Felsenstein, 1985; Grafen, 1989; Pagel and Harvey, 1989; Gittleman and Kot, 1990; Abouheif, 1999; Pagel, 1999; Garland, 2001; Blomberg, Garland and Ives, 2003).

2.3.3 Significant changes in rubisco kinetics occur during the evolution of C_4 photosynthesis

Inspection of the data identified several dependencies in rubisco kinetic traits between C_3 and C_4 plants (Figure 2.2). Specifically, the mean of the distribution of rubisco $S_{C/O}$ values in C_4 species (mean $S_{C/O}$ = 78.7 mol.mol^{-1}) was lower than that observed for rubisco in C_3 species (mean $S_{C/O}$ = 89.9 mol.mol^{-1}) (Figure 2.2; $p < 0.001$, t-test). Conversely, the mean of the distribution of rubisco k_{catC} values was higher in C_4 species (mean k_{catC} = 4.2 s^{-1}) than in C_3 species (mean k_{catC} = 3.2 s^{-1}) (Figure 2.2; $p < 0.001$, t-test). The means of the distributions of both K_C and K_C^{air} were also found to be higher in C_4 species (mean K_C = 19.0 μM, mean K_C^{air} = 29.9 μM) than in C_3 species (mean K_C = 15.4 μM, mean K_C^{air} = 23.6 μM) (Figure 2.2; $p < 0.05$ and $p < 0.05$, t-test, respectively). In contrast, no significant difference was observed in K_O between C_3 species (mean K_O = 481.0 μM) and C_4 species (mean K_O = 466.7 μM) (Figure 2.2; $p > 0.05$, t-test). However, variation in K_O was found to be considerably greater in C_4 species (95 % CI = [379.1, 574.6]) than in C_3 species (95 % CI = [457.1, 506.0]) ($p < 0.01$; Levene test). Further, although the restricted number of K_{RuBP} measurements did not allow statistical differences to be assessed between photosynthetic groups, the distribution of this trait appeared to show higher variability in C_4 species, similar to that observed for K_O. Owing to limited kinetic measurements for rubisco in C_3-C_4 intermediate and C_4-like species which respectively represent early and late transition states along the evolutionary continuum from C_3 to C_4 photosynthesis, it was not possible to assess changes in rubisco kinetics in these plants relative to the ancestral C_3 and derived C_4 photosynthetic types. Nevertheless, trait values of rubisco $S_{C/O}$ in

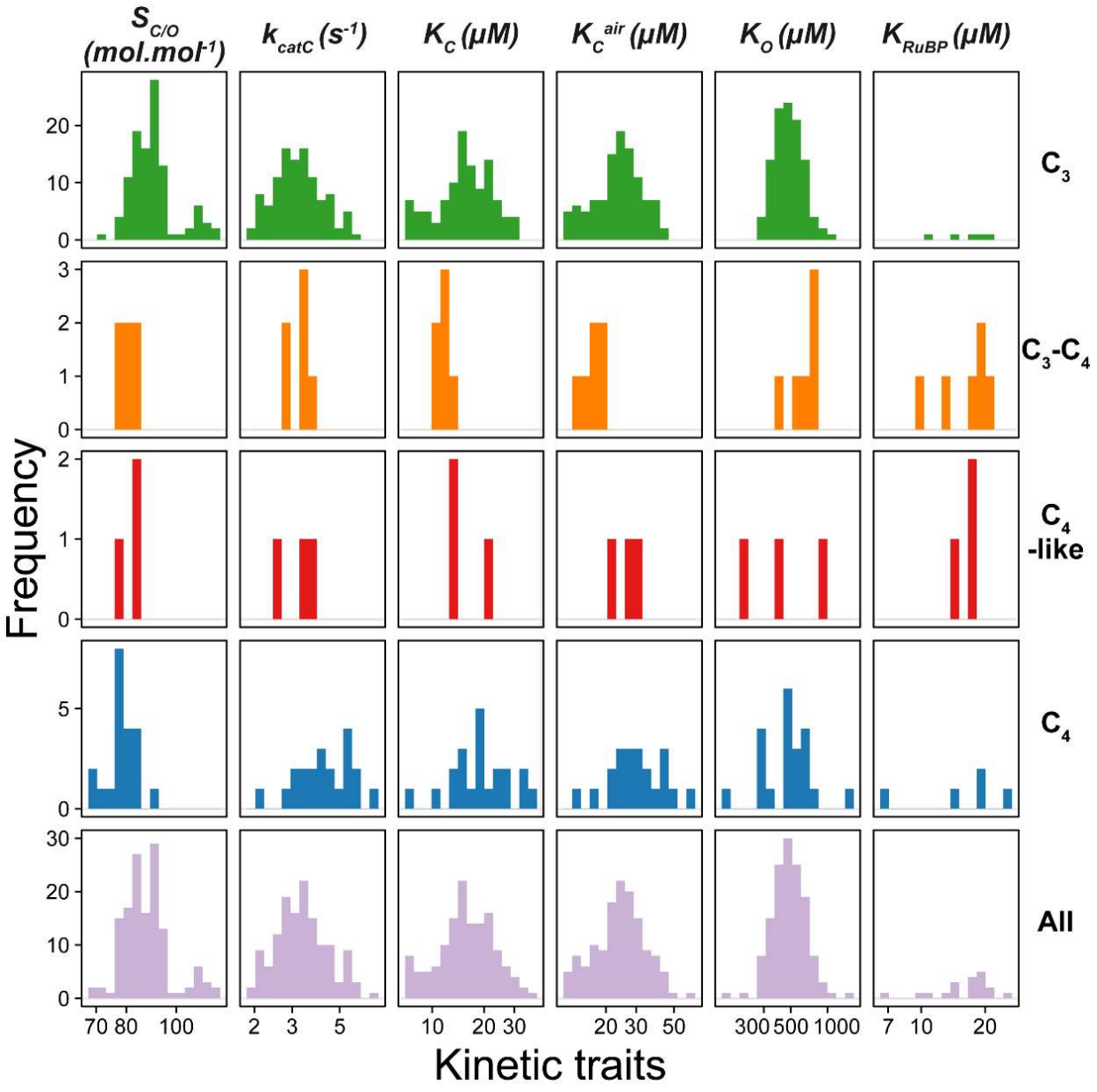


Figure 2.2. The distributions of values for rubisco kinetic traits in angiosperms. Species are grouped by their photosynthetic types (rows). $S_{C/O}$: specificity. k_{catC}: carboxylase turnover per site. K_C: the Michaelis constant for the CO_2. K_C^{air} the inferred Michaelis constant for CO_2 in 20.95 % O_2 air. K_O: the Michaelis constant for O_2. K_{RuBP}: the Michaelis constant for ribulose 1,5-bisphosphate. Plants have been classified as those which perform C_3 photosynthesis (C_3; n = 107), C_4 photosynthesis (C_4; n = 6), C_3-C_4 intermediate (C_3-C_4; n = 3), C_4-like (C_4-like; n = 21). The X axis for all plots is on a log scale, where respective units are shown in column labels. The raw dataset used can be found in ***Supplemental File 2 online***.

both evolutionary intermediate C_3-C_4 and C_4-like states appear to closely resemble the distribution observed in C_4 species (Figure 2.2), thus indicating adaptation of this trait potentially occurred early during the evolution of C_4 photosynthesis. All of the significant differences between C_3 and C_4 plants reported above were robust to correction for the measured phylogenetic signal (Appendix B [and ***Supplemental File 1 online***], table S2). Collectively, these data demonstrate that there are adaptations in rubisco kinetics that are associated with the evolution of C_4 photosynthesis, such that the emergence of the C_4 carbon concentrating mechanism is accompanied by a decreased specificity and CO_2 affinity, and an increased carboxylase turnover.

2.3.4 A significant phylogenetic signal exists in rubisco K_O in C_3 plants

Based on the positions of C_3-C_4 intermediate, C_4-like, and C_4 species in the phylogenetic tree (Appendix B [and ***Supplemental File 1 online***], Figure S1), multiple independent evolutions toward C_4 photosynthesis are present in the dataset. Furthermore, given that transition to C_4 photosynthesis is found above to be associated with adaptive changes in rubisco kinetic traits including a reduction in $S_{C/O}$, an increase in k_{catC}, K_C and K_C^{air}, as well as increased variability in K_O (Figure 2.1; Appendix B [and ***Supplemental File 1 online***], table S2), it was hypothesised that a failure to account for kinetic differences associated with photosynthetic type may have confounded estimations of the phylogenetic signal. For example, kinetic modifications associated with the evolution of C_4 photosynthesis may cause larger differences in rubisco kinetics among closely related C_3 and C_4 species than expected based on evolutionary distance alone. Similarly, the independent evolution of C_4 photosynthesis in distantly related plant lineages could also cause evolutionarily distant species to have similar kinetic trait values by convergence. To evaluate the extent by which these respective issues may have affected quantification of the phylogenetic signal, the above analyses were repeated using only the C_3 angiosperm species present (i.e., with C_3-C_4 intermediate, C_4-like, and C_4 species removed). In general, estimates of the phylogenetic signal in the carboxylase-related kinetic traits in C_3 species (table 2.2) agreed with those observed when all angiosperms were considered on the phylogenetic tree (table 2.1). Specifically, a phylogenetic signal of similar strength and significance was observed in $S_{C/O}$, k_{catC} and K_C for each of the detection methods across both

Table 2.2. The phylogenetic signal strength and associated significance level in rubisco kinetic traits in C_3 species using five different signal detection methods. Statistics are rounded to 3 decimal places, and significance values are represented as α levels, where; α = 0.001 if $p < 0.001$, α = 0.01 if $0.001 < p < 0.01$, α = 0.05 if $0.01 < p < 0.05$, and α = ns if $p > 0.05$.

Kinetic Trait	C mean		I		K		K *		Lambda	
	Stat	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***
$S_{C/O}$	0.533	0.001	0.453	0.001	0	ns	0.001	ns	0.814	0.001
k_{catC}	0.387	0.001	0.234	0.001	0.002	0.01	0.002	0.01	0.913	0.001
K_C	0.449	0.001	0.341	0.001	0.001	0.05	0.001	0.05	0.948	0.001
K_C^{air}	0.398	0.001	0.317	0.001	0.001	0.05	0.002	0.01	0.947	0.001
K_O	0.279	0.001	0.167	0.01	0	ns	0	ns	0.743	0.001

sets of analyses (table 2.1 and 2.2). In addition, the discrepancies in signal strength between the statistical methods previously observed (table 2.1) were recapitulated in the analysis using only C_3 species (table 2.2), thus indicating that these differences are not caused by a failure to control for photosynthetic type, but instead more likely represent distinctions in the assumptions and aspects of the phylogenetic signal measured by each test (Hardy and Pavoine, 2012; Münkemüller *et al.*, 2012). In summary, there is a statistically significant phylogenetic signal in rubisco specificity, carboxylase turnover and the Michaelis constant for CO_2 in angiosperms that is independent of photosynthetic type.

In contrast to the analysis of all angiosperms (table 2.1), a significant signal was observed in K_O when only C_3 angiosperms were considered (table 2.2). Thus, both the oxygenase-related and carboxylase-related traits of rubisco have evolved in a tree-like manner in C_3 angiosperms. Furthermore, unlike the other carboxylase-related kinetic traits, the phylogenetic signal in K_C^{air} was found to increase in strength when the analysis is restricted to C_3 angiosperms. This result is a corollary of the fact that K_C^{air} is computed here from both K_C and K_O. Thus, all kinetic traits of rubisco have a significant phylogenetic signal in C_3 angiosperms.

2.3.5 Correlations between kinetic traits are weak in angiosperms and further relaxed after correcting for a phylogenetic signal

Given the finding that rubisco kinetic traits exhibit a significant phylogenetic signal (table 2.1; Figure 2.2), it is possible that previously reported correlations between rubisco kinetic traits (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010; Flamholz *et al.*, 2019; Iñiguez *et al.*, 2020) are an artefact of this signal. This is because prior analyses which did not account for the inherent phylogenetic structure (and non-independence of) this data (Figure 2.3A) may have suffered from over-estimated correlation coefficients due to this underlying structure. In order to evaluate the severity by which this signal may have influenced previous results (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010; Flamholz *et al.*, 2019; Iñiguez *et al.*, 2020), the correlations observed in the kinetic trait data using both phylogenetic and non-phylogenetic regression methods were compared (Figure 2.3B and 2.3C).

Using a standard non-phylogenetic approach, the relationships between kinetic traits of rubisco were consistent in both linear and least squares regression models (Appendix B [and ***Supplemental File 1 online***], Figure S3A and S3B). The direction of power-law relationships observed across angiosperms (Figure 2.3B) match those previously reported (Flamholz *et al.*, 2019). Specifically, significant positive correlations were found between k_{catC} and both K_C and K_C^{air} (Figure 2.3B and 2.3C). A significant positive correlation was also observed between the respective Michaelis constants for both CO_2 and O_2 substrates, K_C and K_O (Figure 2.3B and 2.3C). In addition, significant inverse power-law correlations were observed between $S_{C/O}$ and all other carboxylase-related kinetic traits investigated, including k_{catC}, K_C and K_C^{air} (Figure 2.3B and 2.3C). In contrast, K_O did not co-vary with either $S_{C/O}$ or k_{catC} (Figure 2.3B and 2.3C), whilst K_{RuBP} did not appear to be tightly linked to any kinetic trait from the limited number of observations that are available (Appendix B [and ***Supplemental File 1 online***], Figure S3A and S3B). Thus, across angiosperms, all pairwise relationships between the carboxylase-related kinetic traits $S_{C/O}$, k_{catC} and either K_C or K_C^{air} were significant, whilst the oxygenase-related trait K_O was only correlated with K_C.

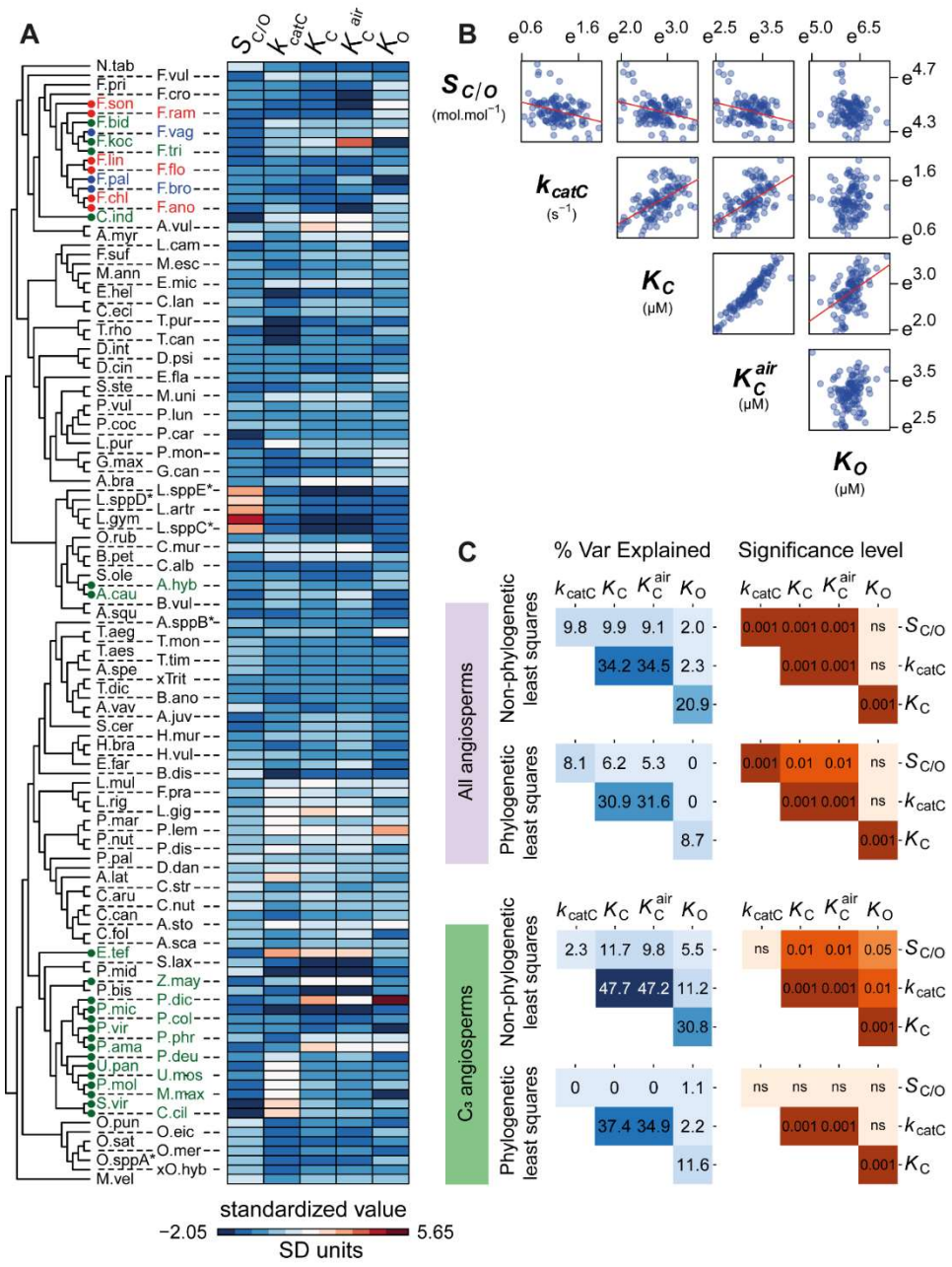


Figure 2.3. The correlations between rubisco kinetic traits in angiosperms. **A**) Heatmap depicting the normalised variation in kinetic traits across the species set used in this study (± S.D. away from each respective kinetic trait mean). Species labels on the tree are colour coded by photosynthetic type (C_3: black, C_3-C_4 intermediates: red, C_4-like: blue, and C_4: green), and have been abbreviated for legibility (refer to Appendix B [and ***Supplemental File 1 online***], table S4). **B)** Trends in relationships between all pairwise combinations of log transformed rubisco kinetic traits shown in A. **C**) Pairwise correlation coefficients (percent variance explained) and associated *p*-values between different rubisco kinetic traits assessed using non-phylogenetic least squares regression models or phylogenetic least squares regression models. Both phylogenetic and non-phylogenetic least squares regression were fit to both the complete set of angiosperms in the dataset and the subset which perform C_3 photosynthesis. Significance values are represented as α levels, where; α = 0.001 if $p < 0.001$, α = 0.01 if $0.001 < p < 0.01$, α = 0.05 if $0.01 < p < 0.05$, and α = ns if $p > 0.05$.

Although kinetic trade-offs inferred using non-phylogenetic methods were concordant in direction with those previously described (Flamholz *et al.*, 2019), they were substantially reduced in magnitude when the analysis was focussed solely on the angiosperms. This reduction in magnitude of correlation when taxonomic groups are removed is strongly indicative of phylogenetic signal in the wider dataset and is analysed in further detail in a subsequent results section. Within angiosperms, the strength of the correlation between $S_{C/O}$ and K_C (9.9 % variance explained; Figure 2.3C) is attenuated by 77 % when compared to that previously reported using a larger range of rubisco variants (43.6 % variance explained (Flamholz *et al.*, 2019)). Moreover, a 69 % reduction was found in the dependency between $S_{C/O}$ and k_{catC} in angiosperms (9.8 % variance explained; Figure 2.3C) in comparison to that reported based on the larger range of species (31.4 % variance explained (Flamholz *et al.*, 2019)), with the antagonistic correlation observed between K_C and K_O (20.9 % variance explained; Figure 2.3C) also weakened by 33 % relative to previous reports (31.4 % variance explained (Flamholz *et al.*, 2019)). In contrast, the dependency between K_C and k_{catC} was 49 % stronger when only angiosperms are assessed, increasing from 23.0 % (Flamholz *et al.*, 2019) to 34.2 % in this study (Figure 2.3C). Therefore overall, even in the absence of correctly accounting for the phylogenetic relationship between rubisco, the apparent catalytic trade-offs observed in angiosperms are weaker than previously thought (Flamholz *et al.*, 2019; Iñiguez *et al.*, 2020).

Given that a significant phylogenetic signal is present in rubisco kinetic traits in angiosperms (table 2.1 and 2.2), a phylogenetic generalized least squares regression analysis (Felsenstein, 1985) was conducted to estimate the magnitude of the catalytic trade-offs accounting for the inherent structure of the data. In comparison to phylogeny-unaware correlations, the phylogenetic regression resulted in a reduction in the majority of kinetic trade-offs (Figure 2.3C). The largest reduction observed was for the correlation between the Michaelis constants K_C and K_O. Here, the correlation was reduced by 58 % (variance explained = 8.7 %) relative to methods which do not correctly account for the non-independence of these measurements (variance explained = 20.9 %; Figure 2.3C). An analogous weak correlation was also observed when the phylogenetic analyses were limited to C_3 species (variance explained = 11.6 %; Figure 2.3C). Thus, changes in K_C have occurred independently of any change on K_O in rubisco since the diversification of the angiosperms.

Phylogenetic correction also resulted in less substantial reductions in the correlation between $S_{C/O}$ and each of the other carboxylase-related traits (Figure 2.3C). Here the dependency between $S_{C/O}$ and k_{catC} was reduced by 18 % from 9.8 to 8.1 %, whilst the dependency between $S_{C/O}$ and both K_C and K_C^{air} were reduced by 37 % and 42 % from 9.9 to 6.2 %, and from 9.1 to 5.3 %, respectively (Figure 2.3C). Furthermore, these correlations were not significant when considering only C_3 species (Figure 2.3C). Thus, during the evolution of rubisco in angiosperms, changes to specificity have occurred with little or no effect on other carboxylase-related kinetic traits, and *vice versa*.

In contrast, the strength of the correlation between k_{catC} and either K_C or K_C^{air} was robust to phylogenetic correction. Specifically, the dependency between k_{catC} and K_C only decreased by 10 % from 34.2 to 30.9 %, and the dependency between k_{catC} and K_C^{air} decreased by only 8 % from 34.5 to 31.6 %; Figure 2.3C). Further, the phylogenetically corrected correlations between these kinetic traits were of a similar magnitude when only C_3 species were considered (37.4 and 34.9 % respectively; Figure 2.3C). Thus, as rubisco kinetics have evolved in angiosperms, there has been a trade-off between CO_2 affinity and carboxylase turnover such that any change in one kinetic trait caused a partial change in the other, though with little impact on any further rubisco kinetic traits.

2.3.6 The evolution of rubisco kinetics is more limited by phylogenetic constraints than by catalytic trade-offs in angiosperms

As rubisco kinetic traits contain a phylogenetic signal in angiosperms (table 2.1 and 2.2), we sought to determine the extent to which the phylogenetic signal was caused by phylogenetic constraint. Here, phylogenetic constraint is considered to comprise all constraints which are embedded within the structure of the phylogenetic tree, that are independent of the kinetic constraints, and collectively act to impede the adaptive evolution of rubisco kinetics. For example, such phylogenetic constraints include processes which lead to neutral evolution (Felsenstein, 1985) or evolutionary stasis (Prinzing *et al.*, 2001; Ackerly, 2004; Cavender-Bares *et al.*, 2004; Moles *et al.*, 2005; Swenson and Enquist, 2007) of the trait in question (Ives, 2019; Nevo *et al.*, 2020). In order to assess the relative strength

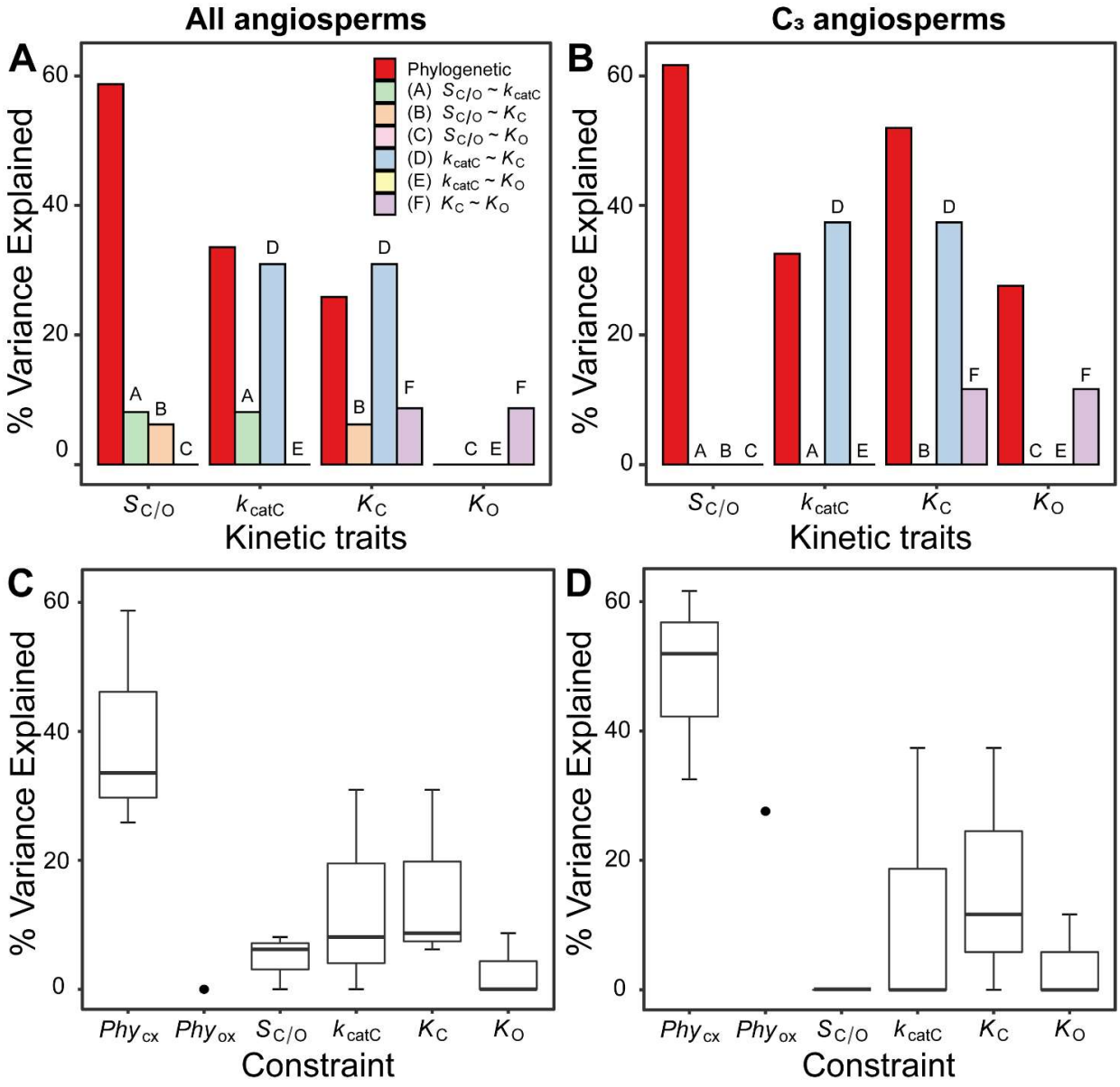


Figure 2.4. The constraints on rubisco kinetic adaptation in angiosperms. **A)** The variation (%) in rubisco kinetic traits across angiosperms that can be explained by phylogenetic constraint and each catalytic trade-off. **B)** As in A but for C_3 angiosperms only. **C)** Boxplot of all variation explained in each kinetic trait by kinetic trait correlations in comparison to variation explained by phylogeny in angiosperms. The phylogenetic constraints on the carboxylase-related traits Phy_{CX} (includes $Phy_{Sc/o}$, Phy_{Kcatc}, and Phy_{Kc}) and phylogenetic constraints on the oxygenase-related trait Phy_{ox} (includes Phy_{Ko} only) are presented separately. **D)** As in C but for C_3 angiosperms only.

of such phylogenetic constraints on rubisco kinetics, the variance in kinetic traits partitioned between phylogenetic effects (i.e., the explanatory power of the phylogenetic tree in the goodness-of-fit model and a measure of phylogenetic constraint) and non-phylogenetic effects (i.e., the remaining explanatory power of the regression model, accounted for by the sum of all other constraints such as random effects and all kinetic trait trade-offs) was quantified. This analysis revealed that phylogenetic constraints explained a significant proportion of variation in carboxylase-related kinetic traits across all angiosperms (Figure 2.4A), and all kinetic traits in C_3 angiosperms (Figure 2.4B). With one exception (i.e. the phylogenetic constraint on K_O in the larger species dataset) the magnitude of variation explained by phylogenetic constraints was similar or larger to the variation explained by the strongest trade-off observed between k_{catC} and K_C (Figure 2.4C and 2.4D). Consequently, across angiosperms the cumulative variance explained by phylogenetic constraints across all rubisco kinetic traits (29.5 %) was larger than the cumulative variance for all catalytic trade-offs combined (9.0 %). This effect was more pronounced for C_3 angiosperms (cumulative variance for phylogenetic constraints = 43.4 %, cumulative variance for catalytic trade-offs = 8.2 %). Thus, during the radiation of angiosperms phylogenetic constraints have restricted the evolution of rubisco kinetics to a greater extent than all catalytic trade-offs combined.

2.3.7 Phylogenetic signal, weak kinetic trait correlations and strong phylogenetic constraint are features of rubisco in all photosynthetic organisms

Given the presence of phylogenetic signal and the impact of phylogenetic constraint on the evolution of rubisco kinetics in angiosperms, we sought to determine whether these findings were a unique feature of angiosperm rubisco or whether they were a more general phenomenon across the tree of life. To achieve this, the dataset was expanded to include all species for which both kinetic measurements and an *rbcL* gene sequence were available. Analogous to the analysis of angiosperms, a strong and statistically significant phylogenetic signal was observed in $S_{C/O}$, k_{catC}, K_C and K_C^{air}, but not in K_O across all photosynthetic organisms (table 2.3). Similarly, a significant phylogenetic signal was also observed for K_O when C_3-C_4, C_4-like and C_4 angiosperms were omitted to control for the dependency in kinetic trait measurements on the tree associated with the convergent transition to C_4 photosynthesis in angiosperms (Appendix B [and ***Supplemental File 1***

Table 2.3. The phylogenetic signal strength and associated significance level in rubisco kinetic traits across all studied photosynthetic organisms using five different signal detection methods. Statistics are rounded to 3 decimal places and significance values are represented as α levels, where; α = 0.001 if $p < 0.001$, α = 0.01 if $0.001 < p < 0.01$, α = 0.05 if $0.01 < p < 0.05$, and α = ns if $p > 0.05$.

Kinetic Trait	C mean		I		K		K *		Lambda	
	Stat	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***
$S_{C/O}$	0.712	0.001	0.229	0.001	0.005	0.001	0.002	0.001	1.005	0.001
k_{catC}	0.538	0.001	0.29	0.001	0.002	0.001	0.001	0.001	0.975	0.001
K_C	0.705	0.001	0.277	0.001	0.002	0.001	0.001	0.001	0.948	0.001
K_C^{air}	0.609	0.001	0.299	0.001	0.001	0.001	0.001	0.01	0.924	0.001
K_O	0.170	0.001	0.004	ns	0	ns	0	ns	0.603	0.001

online], table S3). Thus, there is a significant phylogenetic signal in rubisco kinetic traits in all photosynthetic organisms.

Analogous to above analyses, accounting for the phylogenetic tree (Appendix B [and ***Supplemental File 1 online***], Figure S4) caused a substantial attenuation in the kinetic trait correlations in all species (Figure 2.5A; Appendix B [and ***Supplemental File 1 online***], Figure S3C; Figure S5; Figure S6). Specifically, when correcting for the measured phylogenetic signal in kinetic traits, a partial correlation between k_{catC} and both K_C and K_C^{air} was observed (variance explained = 21.3 % and 23.3 %, respectively; Figure 2.5A). Furthermore, a partial correlation was also measured between K_C and K_O (variance explained = 18.6 %; Figure 2.5A). However, correlations between all other pairwise combinations of kinetic traits were found to be either marginal, or not significant (Figure 2.5A). In addition, the dependency between K_C and K_O was attenuated to 13.4 % when the C_3-C_4, C_4-like and C_4 angiosperms were excluded from the analysis (Appendix B [and ***Supplemental File 1 online***], Figure S7A).

Evaluation of the phylogenetic constraints revealed that they explained a significant proportion of variation in the evolution in all rubisco kinetic traits (Figure 2.5B). Moreover, the phylogenetic constraints explained a larger proportion of kinetic trait variation than catalytic trade-offs (Figure 2.5C), such that the cumulate variation explained by phylogenetic constraints (56.1 %) was larger

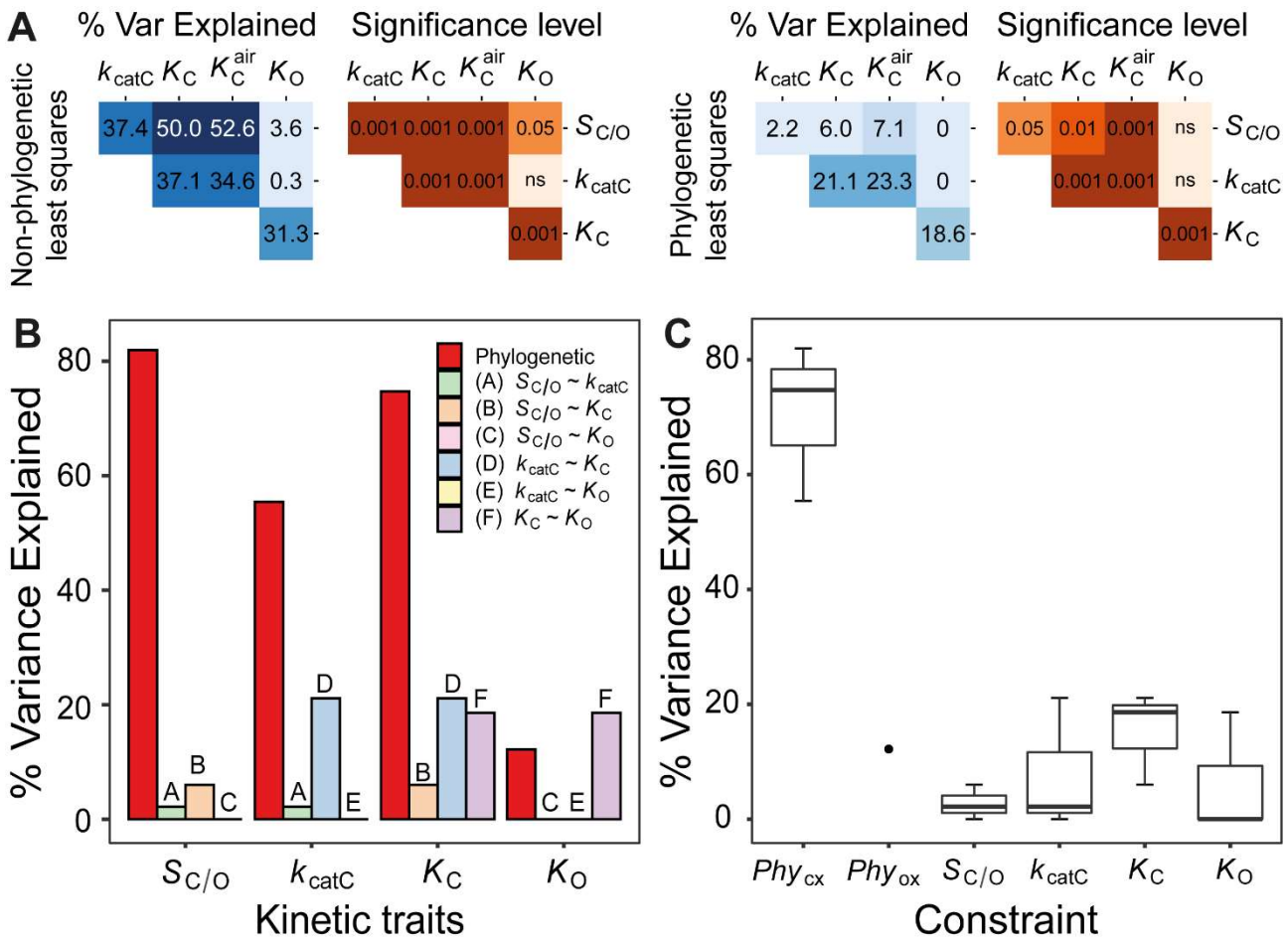


Figure 2.5. Kinetic and phylogenetic constraints on rubisco adaptation across all studied photosynthetic organisms **A)** Pairwise correlation coefficients (percent variance explained) and associated *p*-values between different rubisco kinetic traits assessed using non-phylogenetic least squares regression models or phylogenetic least squares regression models. It should be noted that the non-corrected correlations measured exhibit slight variation with those reported by (Flamholz *et al.*, 2019) owing to the omission of species in this analysis for which *rbcL* sequence data was not available. Significance values are represented as α levels, where; α = 0.001 if $p < 0.001$, α = 0.01 if $0.001 < p < 0.01$, α = 0.05 if $0.01 < p < 0.05$, and α = ns if $p > 0.05$. **B)** The variation (%) in rubisco kinetic traits across all photosynthetic organisms that can be explained by phylogenetic constraint and each catalytic trade-off. **C)** Boxplot of all variation explained in each kinetic trait by kinetic trait correlations in comparison to variation explained by phylogeny in all photosynthetic organisms. The phylogenetic constraints on the carboxylase-related traits Phy_{CX} (includes $Phy_{Sc/o}$, Phy_{Kcatc}, and Phy_{Kc}) and phylogenetic constraints on the oxygenase-related trait Phy_{ox} (includes Phy_{Ko} only) are presented separately.

than the combined effect of all catalytic trade-offs (8.0 %). Analogous results were recovered when C_3-C_4, C_4-like and C_4 species were removed from the analysis (cumulative variance for phylogenetic constraints = 61.4 %, cumulative variance for catalytic trade-offs = 6.2 %; Appendix B [and ***Supplemental File 1 online***], Figure S7B and S7C). Thus, phylogenetic constraints have been a critical limitation on rubisco adaptation in a diverse range of photoautotrophs and have presented a greater barrier to kinetic evolution compared to that imposed by all catalytic trade-offs combined.

2.4 Discussion

The evolutionary landscape of rubisco has long been proposed to be constrained by catalytic trade-offs. In support of this hypothesis, antagonistic correlations between rubisco kinetic traits inferred from studies comparing limited numbers of species are commonly cited (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010). Specifically, strong dependencies are thought to occur between rubisco specificity ($S_{C/O}$), carboxylase turnover (k_{catC}) and the Michaelis constants for CO_2 (K_C) and O_2 (K_O), respectively (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010). Combined, these trade-offs are hypothesized to limit the capacity of rubisco to assimilate CO_2 at high rates by curtailing the inherent carboxylase activity of the enzyme, whilst also causing it to catalyse a reaction with O_2 which is energetically expensive and results in a loss of fixed carbon (Bowes, Ogren and Hageman, 1971; Chollet, 1977). However, all trade-offs have been inferred based on the assumption that rubisco in different species are independent (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010; Flamholz *et al.*, 2019; Iñiguez *et al.*, 2020). Here, we find that this assumption was incorrect and show that a significant phylogenetic signal is found in rubisco kinetic traits across the tree of life, in agreement with (Orr *et al.,* 2016). We re-evaluated the extent of rubisco catalytic trade-offs accounting for this phylogenetic signal and found that all catalytic trade-offs were attenuated. The largest trade-offs were observed between the Michaelis constant for CO_2 and carboxylase turnover (~21-37 %), and between the Michaelis constants for CO_2 and O_2 (~9-19 %), respectively. Furthermore, we demonstrated that all other catalytic trade-offs were either non-significant or substantially attenuated when the phylogenetic relationship of the species was correctly accounted

for. Finally, we found that phylogenetic constraints have played a larger role than catalytic trade-offs in limiting the evolution of rubisco kinetics. Thus, rubisco kinetics have been evolving largely independently of each other in an adaptive landscape that is predominantly limited by phylogenetic constraint.

The presence of a phylogenetic signal in rubisco kinetic traits simply means that rubisco kinetics are more similar among close relatives, with this similarity changing as a function of the phylogenetic distance between species. This result is perhaps not surprising given that all extant rubisco are related by the process of descent with modification from a single ancestral enzyme (Nisbet *et al.*, 2007). However, not all biological traits contain a phylogenetic signal (Blomberg, Garland and Ives, 2003; Kamilar and Cooper, 2013). Further, the functional consequences of changes to enzyme sequences are hard to predict (Minshull *et al.*, 2005; Damborsky and Brezovsky, 2014; Siegel *et al.*, 2015; Carlin *et al.*, 2016), with single amino acid substitutions often causing large effects in enzyme kinetics (Cleton-Jansen *et al.*, 1991; Villar *et al.*, 1997; Johnson *et al.*, 2001; Kapralov *et al.,* 2011; Galmés *et al.,* 2014). Thus, *a priori* it was unknown whether any or all of the rubisco kinetic traits would exhibit a phylogenetic signal. It will be interesting to see whether the presence of phylogenetic signal in enzyme kinetic data is a general phenomenon, and if so it will be important to correctly account for this non-independence when inferring the presence or absence of catalytic trade-offs.

In this work we reveal that the phylogenetic signal in rubisco kinetics is caused by phylogenetic constraint on rubisco that is independent of the catalytic trade-offs. Phylogenetic constraint in this context includes all of the processes that collectively lead to slow rates of molecular evolution. These process include neutral evolution under genetic drift (Felsenstein, 1985), or evolutionary stasis (Prinzing *et al.*, 2001; Ackerly, 2004; Cavender-Bares *et al.*, 2004; Moles *et al.*, 2005; Swenson and Enquist, 2007) under which adaptive change is mitigated by processes including stabilizing selection, pleiotropy and a lack of molecular variability or phenotypic plasticity (Maynard Smith *et al.*, 1985; Bradshaw, 1991; Edwards and Naeem, 1993; Wagner, 1995). Although it is possible that multiple factors contribute to the phylogenetic constraint detected in rubisco, it is likely that low molecular variability is a key driver of this phenomenon. For example, the rate of molecular evolution of rubisco

is likely constrained by the requirements for i) high levels of transcript and protein abundance (Kelly, 2018; Seward and Kelly, 2018), ii) maintaining complementarity to a wide array of molecular chaperones which assist in protein folding and assembly (e.g., Raf1, Raf2, RbcX, BSD2, Cpn60/Cpn20) and metabolic regulation (e.g., rubisco activase) (Carmo-Silva *et al.*, 2015; Aigner *et al.*, 2017), and iii) the need to preserve overall protein stability within the molecular activity-stability trade-offs (Studer *et al.*, 2014; Duraõ *et al.*, 2015; Cummins, Kannappan and Gready, 2018). In angiosperms, these constraints would be further exacerbated due to the presence of the *rbcL* gene in the organellar genome that is uniparentally inherited and does not recombine (Birky, 2001). For example, in angiosperms chloroplast encoded genes evolve 10 times slower than nuclear encoded genes (Wolfe, Li and Sharp, 1987; Smith, 2015). Combined, these evolutionary constraints would hinder the kinetic adaptation of rubisco resulting in the phylogenetic constraint observed in this study.

The strongest catalytic trade-off detected in this study was the ~21-37 % dependency that was observed between k_{catC} and both K_C and K_C^{air}. This finding is compatible with the mechanistic models of rubisco (Farquhar, 1979), and is supported by the recent discovery of enzyme variants which exhibit the highest k_{catC} ever recorded at the expense of poor CO_2 affinities (i.e. K_C values >250 μM) (Davidi *et al.*, 2020). Nevertheless, the dependency between CO_2 affinity and turnover, despite being the strongest correlation that was observed, is substantially attenuated relative to the coefficients that are conventionally cited (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010). Therefore, although selecting for a greater rubisco carboxylase turnover is evolutionarily linked with a poorer affinity for CO_2 (higher K_C), and *vice versa*, significant plasticity exists in this relationship among species such that variation in one kinetic trait only explains ~21-37 % of variation in the other. This fact explains why there is variability in the carboxylation efficiency among angiosperm rubisco (defined as k_{catC}/K_C), a core parameter which defines the initial slope of the response of CO_2 fixation rate to changes in CO_2 concentration within the aerobic environment of chloroplasts in C_3 species (Sharwood, 2017). The second strongest catalytic trade-off that was observed was the ~9-19 % dependency between K_C and K_O. This trade-off is compatible with the fact that the singular active site of rubisco binds both CO_2 and O_2, and thus it is plausible that mutations that affect the active site will affect biding of both substrates, though not necessarily to equivalent extents. All other

catalytic trade-offs were either marginal (<9 %) or non-significant. Furthermore, the combined effect of all catalytic trade-offs can only account for 6 - 9 % of total variation in rubisco kinetics between species, a substantially smaller component than can be explained by phylogenetic constraint (30 - 61%).

The phylogenetically resolved analysis of rubisco kinetic evolution also identified changes in kinetic traits associated with the evolution of C_4 photosynthesis. Specifically, $S_{C/O}$ was lower in C_4 species than in C_3 species, while k_{catC}, K_C and K_C^{air} were higher in C_4 species than in C_3 species. Moreover, variation in K_O was found to be greater in C_4 species than in C_3 species. These differences in rubisco kinetics would likely be either neutral or adaptive in a C_4 context. For example, any change in K_O would effectively be neutral under the elevated CO_2 environment of the bundle sheath chloroplast, as it would have only a marginal effect on the *in vivo* carboxylation rate or carboxylation-to-oxygenation ratio, and thus would not cause a concomitant change to organism fitness. In contrast, an increase in k_{catC} in the same elevated CO_2 environment would enable higher flux through rubisco and thus provide an energetic advantage to the cell. Accordingly, one would expect that an increased variation in K_O in C_4 species would occur by neutral drift (Kimura, 1991; Savir *et al.*, 2010), and that an increased k_{catC} would confer a selective advantage even if it came at the expense of a partial reduction in K_C. Thus, the adaptations to rubisco kinetics that occur concomitant with the evolution of C_4 photosynthesis are consistent with the change in CO_2:O_2 ratio and weak catalytic trade-off that exists between k_{catC} and K_C. Here, despite the phylogenetic constraints limiting rubisco adaptation, the increased rate at which these kinetic changes occurred in C_4 species may have been facilitated by the higher rates of molecular evolution (Kelly, 2018) and diversification (Spriggs, Christin and Edwards, 2014) that occurs concomitant with the evolution of C_4 photosynthesis.

Whilst every effort was taken to prevent systematic or methodological biases from influencing the results, several factors may have led to the underestimation of phylogenetic signal in the data. For example, experimental error in kinetic measurements, and/or inconsistencies associated with measurements being compiled from numerous sources, may have hindered the detection of phylogenetic signal, as has been shown in other studies (Rohlf, 2001). However, to help mitigate this problem, the $S_{C/O}$ values used in this analysis were normalised to avoid the discrepancy between

the rates measured using an oxygen electrode assay (Parry, Keys and Gutteridge, 1989) and those measured using high precision gas-phase-controlled ^{3}H-RuBP-fixation assays (Kane *et al.*, 1994) (see methods). Thus, improvements in both the accuracy and breadth of rubisco kinetic measurements across species will lead to a concomitant improvement in our understanding of how rubisco kinetics have evolved.

Given the importance of rubisco to life on Earth, the question as to why a "perfect" rubisco has not already evolved is legitimate. For example, although rubisco K_C is thought to be near optimal in C_3 plants in light of the ~8 μM chloroplastic concentration of CO_2 and the inherent limitations of CO_2 as a substrate, including its inertness, hydrophobicity and low molecular mass (Andrews and Whitney, 2003; Bar-Even *et al.*, 2011; Bathellier *et al.*, 2018), the observed k_{catC} (~3 s^{-1} per site) has often been considered low (Bar-Even *et al.*, 2011; Tcherkez, 2013; Davidi *et al.*, 2018). In addition, all known rubisco variants catalyse a promiscuous and energetically costly reaction with O_2. However, a recent review of rubisco kinetics relative to those of other enzymes has argued that rubisco is actually not such a bad catalyst (Bathellier *et al.*, 2018). Indeed, the phylogenetically informed analysis of rubisco presented here demonstrates that the kinetic traits have been able to evolve largely independently, with kinetic evolution primarily limited by strong phylogenetic constraint. These constraints induce a lag in adaptive evolution that help to explain why the enzyme is better suited to former environmental conditions.

The study presented here highlights the importance of considering phylogenetic relationships when conducting comparative analyses of enzyme kinetics across species. In doing so, it reveals that rubisco evolution has been only weakly constrained by catalytic trade-offs. Instead, phylogenetic constraints, caused by factors that limit the pace of molecular evolution, have provided a more substantial hindrance to rubisco kinetic evolution. Accordingly, it should be feasible in the current synthetic biology revolution to circumvent this evolutionary barrier on rubisco optimisation. Indeed, promising steps toward this goal have been already demonstrated using directed evolution of the enzyme to generate variants with improved catalytic traits in non-photosynthetic archaea (Wilson, Alonso and Whitney, 2016), photosynthetic bacteria (Zhou and Whitney, 2019), and cyanobacteria

(Wilson *et al.*, 2018). Thus, our findings provide optimism for engineering rubisco in food, fibre and fuel crops to have improved catalytic efficiency.

2.5 Materials and Methods

2.5.1 Kinetic data

Kinetic measurements of rubisco were attained from (Flamholz *et al.*, 2019). $S_{C/O}$ values measured using the O_2 electrode method which calculate [CO_2] using a pKa of 6.11 (Parry, Keys and Gutteridge, 1989) were normalized relative to $S_{C/O}$ values quantified using high precision gas-phase-controlled ^{3}H-RuBP-fixation assays (Kane *et al.*, 1994) in order to minimise methodological biases in the data. Specifically, as rubisco from wheat (*Triticum aestivum*) was measured in both the O_2 electrode studies (Orr *et al.*, 2016; Prins *et al.*, 2016) as well as in the high precision method by (Kane *et al.*, 1994), multipliers were applied to all $S_{C/O}$ measurements derived from O_2 electrode assays using wheat as an enzyme standard. The distribution of $S_{C/O}$ values in angiosperms before and after normalisation can be seen in (Appendix B [and ***Supplemental File 1 online***], Figure S8).

All kinetic traits in the dataset were transformed on a log scale consistent with Flamholz *et al.*, (2019), and the distributional assumptions of each were verified for analyses herein. For the angiosperm focussed analysis, only angiosperms with experimental measurements of all four principal experimentally measured kinetic traits of interest ($S_{C/O}$, k_{catC}, K_C and K_O) were taken forward for subsequent analysis. However, all species in the dataset with greater than one kinetic trait measurement were considered for subsequent analyses of all photosynthetic organisms. Furthermore, a small number of species also had measurements available for $K_{RuBP.}$ In addition, the measure of the Michaelis constant for CO_2 under 20.95 % ambient air, K_C^{air}, was inferred from K_C and K_O based on the formula $K_C + (K_C [O_2] / K_O)$, where 20.95 % [O_2] in water is 253 µM.

In cases where duplicate entries for a species were present in the kinetic dataset (including synonyms), the median value of their respective kinetic traits was used for inference. In this way, averages were also taken between *Triticum timonovum* and *Triticum timopheevii*, the former of which is a synthetic octoploid of the latter (Murashov and Morozova, 2008). The modified dataset

containing corrected $S_{C/O}$ values and averages across duplicate entries for species is provided in ***Supplemental File 2 online***. It should also be noted that values of k_{catC} are presented as units per site.

2.5.2 Phylogenetic tree inference

As sequenced genomes or transcriptomes do not exist for many species in the kinetic trait dataset, whole genome phylogenomic approaches could not be used to infer the species tree necessary in order to detect a phylogenetic signal in measured kinetic traits of the rubisco holoenzyme. However, the *rbcL* gene that encodes the large subunit of rubisco has a long history of use for phylogenetic inference of species relationships (Gielly and Taberlet, 1994; APG, 1998, 2016) and was available for all of the angiosperms, and the majority of photosynthetic organisms, that were considered in the analyses. As such, *rbcL* was selected here for use in species tree inference. The coding sequences of *rbcL* for the 137 angiosperm species for which kinetic data was available can be found in ***Supplemental File 3 online***. The coding sequences of *rbcL* for the complete set of 181 photosynthetic organisms for which both kinetic data and sequencing data was available can be found in ***Supplemental File 4 online***. Gene sequences were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/) for all species except *Flaveria brownii* which was acquired from the 1KP database (Leebens-Mack *et al.*, 2019). Multiple sequence alignments were performed using mafft L-INS-i (Katoh and Standley, 2013), and alignments were trimmed at the terminal ends to remove unaligned positions using AliView (Larsson, 2014). These trimmed nucleotide sequence alignments were used for subsequent phylogenetic analysis. Bootstrapped maximum likelihood phylogenetic trees were inferred by IQ-TREE (Nguyen *et al.*, 2015) using the ultrafast bootstrapping method with 1000 replicates and the Shimodaira-Hasegawa approximate likelihood ratio branch test. The best fitting model of sequence evolution was inferred from the data automatically by IQTREE. The resultant maximum likelihood phylogenetic trees were rooted manually using Dendroscope (Huson and Scornavacca, 2012). A number of nodes in the angiosperm tree (Appendix B [and ***Supplemental File 1 online***], Figure S1) and a number of nodes in the tree of all photosynthetic organisms exhibited terminal zero-length branches due to 100 % sequence identify with other closely related species (n = 18 and n = 23, respectively). These species were condensed into single data

points (as their *rbcL* are 100 % identical) and the mean of their kinetic traits was used. This reduced the dataset to 119 unique angiosperm enzymes and 158 unique enzymes across all photosynthetic organisms. The phylogenetic tree inferred from the angiosperm taxa (Appendix B [and ***Supplemental File 1 online***], Figure S1) closely matched the topology of the phylogenetic tree expected from the angiosperm phylogeny with only a few alterations (APG, 2016). Moreover, the topology of the *rbcL* trees most accurately reflects the sequence similarity of the rubisco genes and thus were deemed as suitable for investigation of any phylogenetic signal and its effects on correlations between rubisco kinetic traits.

To confirm that the phylogenetic signal was not attributable to overfitting caused by the use of the *rbcL* gene to infer the phylogeny of rubisco, an analogous maximum likelihood angiosperm phylogenetic tree was inferred using IQ-TREE (Nguyen *et al.*, 2015) following the methods described above but based on the angiosperm multiple sequence alignment in which codons containing non-synonymous nucleotide sequence changes were removed (***Supplemental File 5 online***). Due to the considerable loss of phylogenetic information accessible for tree building from this alignment, the species tree inferred using nucleotide sequences corresponding to ubiquitously conserved amino acid positions (Appendix B [and ***Supplemental File 1 online***], Figure S2) exhibited an additional number (n = 13) of angiosperm species belonging to terminal zero-length branches. As the sequences of these species are known to exhibit non-synonymous mutations which are not included in the tree, it is not appropriate to take means of their kinetic traits as above. As such, these data points were removed from analysis using only this tree. Use of this phylogenetic tree confirmed that the presence of the phylogenetic signal in kinetic traits was not due to overfitting, however as this tree was less accurate than the full-length alignment tree, it was not used for any subsequent analysis.

2.5.3 Phylogenetic signal analysis

The presence of a phylogenetic signal in kinetic traits was assessed using five different phylogenetic signal detection methods (Gittleman and Kot, 1990; Abouheif, 1999; Pagel, 1999; Blomberg, Garland and Ives, 2003). Here, signal strength was estimated by assessing the distribution of trait values relative to the respective underlying species tree inferred from the *rbcL* sequences using methods

which both depend on an explicit evolutionary model, such as Pagel's lambda (Pagel, 1999) and Blomberg's K and K* (Blomberg, Garland and Ives, 2003), as well as the spatial autocorrelation analyses of Moran's I (Gittleman and Kot, 1990) and Abouheif's Cmean (Abouheif, 1999). Implementation of these phylogenetic signal detection tools was performed using the *phyloSignal* function of the phylosignal package (Keck *et al.*, 2016) in the R environment. For further discussion of the differences between phylogenetic signal detection methods see (Münkemüller *et al.*, 2012).

2.5.4 Ancestral state estimation and mapping of kinetic traits to the phylogenetic tree

Ancestral state estimation was conducted to visualise the evolution of rubisco kinetic traits on the phylogenetic tree which relates the angiosperms. For this purpose, the kinetic dataset was mapped and scaled onto the angiosperm species tree by employing the function *ggtree* in the ggtree package (Yu *et al.*, 2017). Here, terminal branches were coloured according to the measurement of the kinetic trait in the species which comprise the terminal branch, whereas internal branches were coloured based on values inferred in ancestral species using ancestral state estimation (Yu *et al.*, 2017).

2.5.5 Least squares and linear regression models

All regression models between pairwise combinations of kinetic traits were fit in the R environment. Phylogenetic generalized regression accounting for the phylogenetic non-independence between species was performed using the function *pgls* in the caper package (Comparative Analyses of Phylogenetics and Evolution in R) (Orme *et al.*, 2014). In each case, the structure of the phylogenetic signal was corrected by using branch length transformations of the phylogenetic tree based on the mean maximum likelihood estimates of lambda calculated for the traits being interrogated, with kapa and delta held constant. In cases where the mean maximum likelihood estimates of lambda exceeded the upper limit of the model, this value was set to 1. Phylogenetic corrections to differences in kinetic trait values between C_3 and C_4 plants based on the phylogenetic non-independence of species were also applied using the *pgls* function in the caper package (Orme *et al.*, 2014) with photosynthetic type incorporated as a factorial variable.

In order to partition the variance in rubisco kinetic traits explained by phylogenetic constraints as compared to non-phylogenetic constraints, the rr2 package (Ives and Li, 2018) was employed in R.

Here, to assess the contribution of phylogeny in explaining the variation in kinetic trait values, the explanatory power of the phylogenetic component was measured by comparing full and reduced phylogenetic regression models using the partial R^2_{pred} inferential statistical, based on advice from (Ives, 2019) of this being the most intuitive and direct approach to understand how different model components explain variation. For this analysis, phylogenetic regression models were fit using the *phylolm* function in the phylolm package (Tung Ho and Ané, 2014) using Pagel's lambda model for the error term.

2.6 References

Abouheif, E. (1999) A method for testing the assumption of phylogenetic independence in comparative data, Evolutionary Ecology Research. doi: 10.1007/s11295-009-0238-5.

Ackerly, D. D. (2004) 'Adaptation, niche conservatism, and convergence: Comparative studies of leaf evolution in the California chaparral', American Naturalist, 163(5), pp. 654–671. doi: 10.1086/383062.

Aigner, H. et al. (2017) 'Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2', Science. doi: 10.1126/science.aap9221.

Andersson, I. and Backlund, A. (2008) 'Structure and function of Rubisco', Plant Physiology and Biochemistry, 46(3), pp. 275–291. doi: 10.1016/j.plaphy.2008.01.001.

Andrews, T. J. (1988) 'Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits.', The Journal of biological chemistry.

Andrews, T. J. and Whitney, S. M. (2003) 'Manipulating ribulose bisphosphate carboxylase/oxygenase in the chloroplasts of higher plants', Archives of Biochemistry and Biophysics, 414(2), pp. 159–169. doi: 10.1016/S0003-9861(03)00100-0.

APG (1998) An Ordinal Classification for the Families of Flowering Plants, Annals of the Missouri Botanical Garden. doi: 10.2307/2992015.

APG (2016) 'An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV', Botanical Journal of the Linnean Society. Oxford University Press, 181(1), pp. 1–20. doi: 10.1111/boj.12385.

Araújo, W. L. et al. (2017) 'Engineering photosynthesis: Progress and perspectives', F1000Research. doi: 10.12688/f1000research.12181.1.

Badger, M. R. and Lorimer, G. H. (1976) 'Activation of ribulose-1,5-bisphosphate oxygenase. The role of Mg2+, CO2, and pH', Archives of Biochemistry and Biophysics, pp. 723–729. doi: 10.1016/0003-9861(76)90565-8.

Banda, D. M. et al. (2020) 'Novel bacterial clade reveals origin of form I Rubisco', Nature Plants. doi: 10.1038/s41477-020-00762-4.

Bar-Even, A. et al. (2011) ‘The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters’, Biochemistry, 50(21), pp. 4402–4410. doi: 10.1021/bi2002289.

Bar-On, Y. M. and Milo, R. (2019) ‘The global mass and average rate of rubisco’, Proceedings of the National Academy of Sciences of the United States of America. PNAS, 116(10), pp. 4738–4743. doi: 10.1073/pnas.1816654116.

Bathellier, C. et al. (2018) ‘Rubisco is not really so bad’, Plant Cell and Environment, pp. 705–716. doi: 10.1111/pce.13149.

Bauwe, H. et al. (2012) ‘Photorespiration has a dual origin and manifold links to central metabolism’, Current Opinion in Plant Biology, pp. 269–275. doi: 10.1016/j.pbi.2012.01.008.

Beer, C. et al. (2010) ‘Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate’, Science, 329(5993), pp. 834–838. doi: 10.1126/science.1184984.

Birky, C. W. (2001) ‘The Inheritance of Genes in Mitochondria and Chloroplasts: Laws, Mechanisms, and Models’, Annual Review of Genetics. doi: 10.1146/annurev.genet.35.102401.090231.

Blomberg, S. P., Garland, T. and Ives, A. R. (2003) ‘Testing for phylogenetic signal in comparative data: Behavioral traits are more labile’, Evolution. Oxford, UK: Blackwell Publishing Ltd, pp. 717–745. doi: 10.1111/j.0014-3820.2003.tb00285.x.

Bowes, G., Ogren, W. L. and Hageman, R. H. (1971) Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase, Biochemical and Biophysical Research Communications. doi: 10.1016/0006-291X(71)90475-X.

Bradshaw, A. (1991) ‘Genostasis and the limits to evolution’, Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 333(1267), pp. 289–305.

Busch, F. A. (2020) ‘Photorespiration in the context of Rubisco biochemistry, CO2 diffusion and metabolism’, Plant Journal, 101(4), pp. 919–939. doi: 10.1111/tpj.14674.

Carlin, D. A. et al. (2016) ‘Kinetic characterization of 100 glycoside hydrolase mutants enables the discovery of structural features correlated with kinetic constants’, PLoS ONE. doi: 10.1371/journal.pone.0147596.

Carmo-Silva, E. et al. (2015) ‘Optimizing Rubisco and its regulation for greater resource use efficiency’, Plant, Cell and Environment, pp. 1817–1832. doi: 10.1111/pce.12425.

Cavender-Bares, J. et al. (2004) ‘Phylogenetic overdispersion in Floridian oak communities’, American Naturalist, 163(6), pp. 823–843. doi: 10.1086/386375.

Cheverud, J. M., Dow, M. M. and Leutenegger, W. (1985) ‘the Quantitative Assessment of Phylogenetic Constraints in Comparative Analyses: Sexual Dimorphism in Body Weight Among Primates’, Evolution. Wiley, 39(6), pp. 1335–1351. doi: 10.1111/j.1558-5646.1985.tb05699.x.

Cheverud, J. M., Dow, M. M. and Leutenegger, W. (1986) ‘A Phylogenetic Autocorrelation Analysis of Sexual Dimorphism in Primates’, American Anthropologist. Wiley-Blackwell, 88(4), pp. 916–922. doi: 10.1525/aa.1986.88.4.02a00090.

Chollet, R. (1977) The biochemistry of photorespiration, Trends in Biochemical Sciences. doi: 10.1016/0968-0004(77)90364-4.

Chollet, R. and Anderson, L. L. (1976) ‘Regulation of ribulose 1,5-bisphosphate carboxylase-oxygenase activities by temperature pretreatment and chloroplast metabolites’, Archives of Biochemistry and Biophysics, pp. 344–351. doi: 10.1016/0003-9861(76)90173-9.

Cleton-Jansen, A. M. et al. (1991) ‘A single amino acid substitution changes the substrate specificity of quinoprotein glucose dehydrogenase in Gluconobacter oxydans’, MGG Molecular & General Genetics. doi: 10.1007/BF00272157.

Cummins, P. L., Kannappan, B. and Gready, J. E. (2018) ‘Directions for optimization of photosynthetic carbon fixation: Rubisco’s efficiency may not be so constrained after all’, Frontiers in Plant Science, 9. doi: 10.3389/fpls.2018.00183.

Damborsky, J. and Brezovsky, J. (2014) ‘Computational tools for designing and engineering enzymes’, Current Opinion in Chemical Biology. doi: 10.1016/j.cbpa.2013.12.003.

Davidi, D. et al. (2018) ‘A Bird’s-Eye View of Enzyme Evolution: Chemical, Physicochemical, and Physiological Considerations’, Chemical Reviews, 118(18), pp. 8786–8797. doi: 10.1021/acs.chemrev.8b00039.

Davidi, D. et al. (2020) ‘Highly active rubiscos discovered by systematic interrogation of natural sequence diversity’, The EMBO Journal. John Wiley & Sons, Ltd. doi: 10.15252/embj.2019104081.

Duraõ, P. et al. (2015) ‘Opposing effects of folding and assembly chaperones on evolvability of Rubisco’, Nature Chemical Biology, 11(2), pp. 148–155. doi: 10.1038/nchembio.1715.

Eckardt, N. A. (2005) ‘Photorespiration revisited’, Plant Cell. American Society of Plant Biologists, pp. 2139–2141. doi: 10.1105/tpc.105.035873.

Edwards, E. J. (2019) ‘Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis’, New Phytologist, pp. 1742–1755. doi: 10.1111/nph.15851.

Edwards, S. V. and Naeem, S. (1993) ‘The phylogenetic component of cooperative breeding in perching birds’, American Naturalist. Am Nat, 141(5), pp. 754–789. doi: 10.1086/285504.

Ehleringer, J. R. et al. (1991) Climate change and the evolution of C4 photosynthesis, Trends in Ecology and Evolution. doi: 10.1016/0169-5347(91)90183-X.

Ellis, R. J. (1979) The most abundant protein in the world, Trends in Biochemical Sciences. doi: 10.1016/0968-0004(79)90212-3.

Erb, T. J. and Zarzycki, J. (2018) A short history of RubisCO: the rise and fall (?) of Nature’s predominant CO2 fixing enzyme, Current Opinion in Biotechnology. doi: 10.1016/j.copbio.2017.07.017.

Espie, G. S. and Kimber, M. S. (2011) ‘Carboxysomes: Cyanobacterial RubisCO comes in small packages’, in Photosynthesis Research. doi: 10.1007/s11120-011-9656-y.

Farquhar, G. D. (1979) ‘Models describing the kinetics of ribulose biphosphate carboxylase-oxygenase’, Archives of Biochemistry and Biophysics, 193(2), pp. 456–468. doi: 10.1016/0003-9861(79)90052-3.

Felsenstein, J. (1985) Phylogenies and the comparative method., American Naturalist. doi: 10.1086/284325.

Flamholz, A. I. et al. (2019) ‘Revisiting Trade-offs between Rubisco Kinetic Parameters’, Biochemistry, pp. 3365–3376. doi: 10.1021/acs.biochem.9b00237.

Flamholz, A. and Shih, P. M. (2020) ‘Cell biology of photosynthesis over geologic time’, Current Biology. doi: 10.1016/j.cub.2020.01.076.

Fukayama, H. et al. (2019) ‘Rubisco small subunits of C4 plants, Napier grass and guinea grass confer C4-like catalytic properties on Rubisco in rice’, Plant Production Science. doi: 10.1080/1343943X.2018.1540279.

Galmés, J., Andralojc, P.J., et al. (2014) ‘Environmentally driven evolution of Rubisco and improved photosynthesis and growth within the C3 genus Limonium (Plumbaginaceae)’, New Phytologist, 203(3), pp. 989–999. Available at: https://doi.org/10.1111/nph.12858.

Garland, T. (2001) ‘Phylogenetic comparison and artificial selection’, in. Springer, Boston, MA, pp. 107–132. doi: 10.1007/978-1-4757-3401-0_9.

Gielly, L. and Taberlet, P. (1994) The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences., Molecular Biology and Evolution. doi: 10.1093/oxfordjournals.molbev.a040157.

Gittleman, J. L. and Kot, M. (1990) ‘Adaptation: Statistics and a Null Model for Estimating Phylogenetic Effects’, Systematic Zoology. Society of Systematic Zoology, pp. 227–241. doi: 10.2307/2992183.

Grafen, A. (1989) ‘The phylogenetic regression.’, Philosophical transactions of the Royal Society of London. Series B, Biological sciences. doi: 10.1098/rstb.1989.0106.

Gunn, L. H. et al. (2020) ‘The dependency of red Rubisco on its cognate activase for enhancing plant photosynthesis and growth’, Proceedings of the National Academy of Sciences of the United States of America. doi: 10.1073/pnas.2011641117.

Hardy, O. J. and Pavoine, S. (2012) ‘Assessing phylogenetic signal with measurement error: A comparison of mantel tests, blomberg et al.’s K, and phylogenetic distograms’, Evolution. John Wiley & Sons, Ltd, 66(8), pp. 2614–2621. doi: 10.1111/j.1558-5646.2012.01623.x.

Hatch, M. D. (1987) ‘C4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure’, BBA Reviews On Bioenergetics. doi: 10.1016/S0304-4173(87)80009-5.

Huson, D. H. and Scornavacca, C. (2012) ‘Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks’, Systematic Biology. Oxford University Press, pp. 1061–1067. doi: 10.1093/sysbio/sys062.

Iñiguez, C. et al. (2020) ‘Evolutionary trends in RuBisCO kinetics and their co-evolution with CO2 concentrating mechanisms’, Plant Journal, 101(4), pp. 897–918. doi: 10.1111/tpj.14643.

Ishikawa, C. et al. (2011) ‘Functional incorporation of sorghum small subunit increases the catalytic turnover rate of rubisco in transgenic rice’, Plant Physiology. doi: 10.1104/pp.111.177030.

Ives, A. and Li, D. (2018) ‘rr2: An R package to calculate R^2s for regression models’, Journal of Open Source Software, 3(30), p. 1028. doi: 10.21105/joss.01028.

Ives, A. R. (2019) ‘R 2 s for Correlated Data: Phylogenetic Models, LMMs, and GLMMs’, Systematic Biology, 68(2), pp. 234–251. doi: 10.1093/sysbio/syy060.

Johnson, E. T. et al. (2001) ‘Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase’, Plant Journal. doi: 10.1046/j.1365-313X.2001.00962.x.

Joshi, J. et al. (2015) ‘Role of small subunit in mediating assembly of red-type Form I Rubisco’, Journal of Biological Chemistry. doi: 10.1074/jbc.M114.613091.

Kamilar, J. M. and Cooper, N. (2013) 'Phylogenetic signal in primate behaviour, ecology and life history', Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1618). doi: 10.1098/rstb.2012.0341.

Kane, H. J. et al. (1994) 'An improved method for measuring the CO2/O2 specificity of ribulosebisphosphatte carboxylase-oxygenase', Australian Journal of Plant Physiology. doi: 10.1071/PP9940449.

Kaplan, A. et al. (1991) 'Physiological and molecular aspects of the inorganic carbon-concentrating mechanism in cyanobacteria', Plant Physiology. doi: 10.1104/pp.97.3.851.

Kapralov, M. V. et al. (2011) 'Changes in Rubisco kinetics during the evolution of C4 Photosynthesis in Flaveria (Asteraceae) are associated with positive selection on genes encoding the enzyme', Molecular Biology and Evolution, 28(4), pp. 1491–1503. Available at: https://doi.org/10.1093/molbev/msq335

Katoh, K. and Standley, D. M. (2013) 'MAFFT multiple sequence alignment software version 7: Improvements in performance and usability', Molecular Biology and Evolution. Oxford University Press, pp. 772–780. doi: 10.1093/molbev/mst010.

Keck, F. et al. (2016) 'Phylosignal: An R package to measure, test, and explore the phylogenetic signal', Ecology and Evolution, pp. 2774–2780. doi: 10.1002/ece3.2051.

Kelly, S. (2018) 'The amount of nitrogen used for photosynthesis modulates molecular evolution in plants', Molecular Biology and Evolution, 35(7), pp. 1616–1625. doi: 10.1093/molbev/msy043.

Kimura, M. (1991) 'The neutral theory of molecular evolution: A review of recent evidence', The Japanese Journal of Genetics. doi: 10.1266/jjg.66.367.

Larsson, A. (2014) 'AliView: A fast and lightweight alignment viewer and editor for large datasets', Bioinformatics. Oxford University Press, pp. 3276–3278. doi: 10.1093/bioinformatics/btu531.

Lee, B., Berka, R. M. and Tabita, F. R. (1991) 'Mutations in the small subunit of cyanobacterial ribulose-bisphosphate carboxylase/oxygenase that modulate interactions with large subunits', Journal of Biological Chemistry.

Lee, B. and Tabita, F. R. (1990) 'Purification of Recombinant Ribulose-l,5-bisphosphate Carboxylase/Oxygenase Large Subunits Suitable for Reconstitution and Assembly of Active L8S8 Enzyme', Biochemistry. doi: 10.1021/bi00492a007.

Leebens-Mack, J. H. et al. (2019) 'One thousand plant transcriptomes and the phylogenomics of green plants', Nature. doi: 10.1038/s41586-019-1693-2.

Martin-Avila, E. et al. (2020) 'Modifying Plant Photosynthesis and Growth via Simultaneous Chloroplast Transformation of Rubisco Large and Small Subunits.', The Plant cell. doi: 10.1105/tpc.20.00288.

Maynard Smith, J. et al. (1985) 'Developmental constraints on evolution', The Quarterly Review of Biology.

Meyer, M. and Griffiths, H. (2013) 'Origins and diversity of eukaryotic CO2-concentrating mechanisms: Lessons for the future', Journal of Experimental Botany, pp. 769–786. doi: 10.1093/jxb/ers390.

Minshull, J. et al. (2005) 'Predicting enzyme function from protein sequence', Current Opinion in Chemical Biology. doi: 10.1016/j.cbpa.2005.02.003.

Moles, A. T. et al. (2005) 'A brief history of seed size', Science, 307(5709), pp. 576–580. doi: 10.1126/science.1104863.

Münkemüller, T. et al. (2012) 'How to measure and test phylogenetic signal', Methods in Ecology and Evolution, 3(4), pp. 743–756. doi: 10.1111/j.2041-210X.2012.00196.x.

Murashov, V. V. and Morozova, Z. A. (2008) 'Comparative morphogenesis of Triticum timopheevii (Zhuk.) and synthetic octoploid species T. timonovum Heslot et Ferrary', Moscow University Biological Sciences Bulletin. Heidelberg: Allerton Press, Inc., pp. 127–133. doi: 10.3103/s0096392508030073.

Nevo, O. et al. (2020) 'The evolution of fruit scent: phylogenetic and developmental constraints', BMC Evolutionary Biology. doi: 10.1186/s12862-020-01708-2.

Nguyen, L. T. et al. (2015) 'IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies', Molecular Biology and Evolution, pp. 268–274. doi: 10.1093/molbev/msu300.

Nisbet, E. G. et al. (2007) 'The age of Rubisco: The evolution of oxygenic photosynthesis', Geobiology, 5(4), pp. 311–335. doi: 10.1111/j.1472-4669.2007.00127.x.

Ogren, W. L. (1984) 'Photorespiration: Pathways, Regulation, and Modification', Annual Review of Plant Physiology. Annual Reviews, 35(1), pp. 415–442. doi: 10.1146/annurev.pp.35.060184.002215.

Ogren, W. L. and Bowes, G. (1971) 'Ribulose diphosphate carboxylase regulates soybean photorespiration', Nature New Biology. Nat New Biol, 230(13), pp. 159–160. doi: 10.1038/newbio230159a0.

Orme, D. et al. (2014) Caper: Comparative analyses of phylogenetics and evolution in R, R package version 0.5.2/ r121. Available at: http://r-forge.r-project.org/projects/ caper/ (Accessed: 31 May 2020).

Orr, D. J. et al. (2016) 'Surveying rubisco diversity and temperature response to improve crop photosynthetic efficiency', Plant Physiology. doi: 10.1104/pp.16.00750.

Pagel, M. (1999) 'Inferring the historical patterns of biological evolution', Nature. Nature Publishing Group, pp. 877–884. doi: 10.1038/44766.

Pagel, M. D. and Harvey, P. H. (1989) 'Comparative methods for examining adaptation depend on evolutionary models.', Folia primatologica; international journal of primatology. doi: 10.1159/000156417.

Parry, M. A. et al. (2013) 'Rubisco activity and regulation as targets for crop improvement | Journal of Experimental Botany | Oxford Academic', Journal of Experimental Bitany.

Parry, M. A. J. et al. (2007) 'Prospects for increasing photosynthesis by overcoming the limitations of Rubisco', in Journal of Agricultural Science. doi: 10.1017/S0021859606006666.

Parry, M. A. J., Keys, A. J. and Gutteridge, S. (1989) 'Variation in the specificity factor of C3 higher plant rubiscos determined by the total consumption of ribulose-P2', Journal of Experimental Botany. doi: 10.1093/jxb/40.3.317.

Peterhansel, C. et al. (2010) 'Photorespiration', The Arabidopsis Book. BioOne, 8(8), p. e0130. doi: 10.1199/tab.0130.

Prins, A. et al. (2016) ‘Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis’, Journal of Experimental Botany. doi: 10.1093/jxb/erv574.

Prinzing, A. et al. (2001) ‘The niche of higher plants: Evidence for phylogenetic conservatism’, Proceedings of the Royal Society B: Biological Sciences, 268(1483), pp. 2383–2389. doi: 10.1098/rspb.2001.1801.

Read, B. A. and Tabita, F. R. (1992a) ‘A Hybrid Ribulosebisphosphate Carboxylase/Oxygenase Enzyme Exhibiting a Substantial Increase in Substrate Specificity Factor’, Biochemistry. doi: 10.1021/bi00139a018.

Read, B. A. and Tabita, F. R. (1992b) ‘Amino Acid Substitutions in the Small Subunit of Ribulose-1, 5-bisphosphate Carboxylase/Oxygenase That Influence Catalytic Activity of the Holoenzyme’, Biochemistry. doi: 10.1021/bi00117a031.

Rohlf, F. J. (2001) Comparative methods for the analysis of continuous variables: Geometric interpretations, Evolution. doi: 10.1111/j.0014-3820.2001.tb00731.x.

Sage, R. F., Sage, T. L. and Kocacinar, F. (2012) ‘ Photorespiration and the Evolution of C 4 Photosynthesis ’, Annual Review of Plant Biology, pp. 19–47. doi: 10.1146/annurev-arplant-042811-105511.

Savir, Y. et al. (2010) ‘Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape’, Proceedings of the National Academy of Sciences of the United States of America, 107(8), pp. 3475–3480. doi: 10.1073/pnas.0911663107.

Schneider, G., Lindqvist, Y. and Brändén, C. I. (1992) ‘RUBISCO: Structure and mechanism’, Annual Review of Biophysics and Biomolecular Structure. doi: 10.1146/annurev.bb.21.060192.001003.

Seward, E. A. and Kelly, S. (2018) ‘Selection-driven cost-efficiency optimization of transcripts modulates gene evolutionary rate in bacteria’, Genome Biology, 19(1). doi: 10.1186/s13059-018-1480-7.

Sharkey, T. D. (2020) ‘Emerging research in plant photosynthesis’, Emerging Topics in Life Sciences. Portland Press Ltd. doi: 10.1042/etls20200035.

Sharwood, R. E. (2017) ‘Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops’, New Phytologist, pp. 494–510. doi: 10.1111/nph.14351.

Sharwood, R. E., Ghannoum, O. and Whitney, S. M. (2016) ‘Prospects for improving CO2 fixation in C3-crops through understanding C4-Rubisco biogenesis and catalytic diversity’, Current Opinion in Plant Biology. doi: 10.1016/j.pbi.2016.04.002.

Siegel, J. B. et al. (2015) ‘Computational protein design enables a novel one-carbon assimilation pathway’, Proceedings of the National Academy of Sciences of the United States of America. doi: 10.1073/pnas.1500545112.

Smith, D. R. (2015) ‘Mutation rates in plastid genomes: They are lower than you might think’, Genome Biology and Evolution. doi: 10.1093/gbe/evv069.

Spreitzer, R. J., Peddi, S. R. and Satagopan, S. (2005) ‘Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco’, Proceedings of the National Academy of Sciences of the United States of America. doi: 10.1073/pnas.0508042102.

Spriggs, E. L., Christin, P. A. and Edwards, E. J. (2014) 'C4 photosynthesis promoted species diversification during the miocene grassland expansion', PLoS ONE. doi: 10.1371/journal.pone.0097722.

Studer, R. A. et al. (2014) 'Stability-activity tradeoffs constrain the adaptive evolution of RubisCO', Proceedings of the National Academy of Sciences of the United States of America, 111(6), pp. 2223–2228. doi: 10.1073/pnas.1310811111.

Su, Z. and Townsend, J. P. (2015) 'Utility of characters evolving at diverse rates of evolution to resolve quartet trees with unequal branch lengths: Analytical predictions of long-branch effects', BMC Evolutionary Biology, p. 86. doi: 10.1186/s12862-015-0364-7.

Swenson, N. G. and Enquist, B. J. (2007) 'Ecological and evolutionary determinants of a key plant functional trait: Wood density and its community-wide variation across latitude and elevation', American Journal of Botany, 94(3), pp. 451–459. doi: 10.3732/ajb.94.3.451.

Tabita, F. R. et al. (2008) 'Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships', Journal of Experimental Botany, 59(7), pp. 1515–1524. doi: 10.1093/jxb/erm361.

Tcherkez, G. (2013) 'Modelling the reaction mechanism of ribulose-1,5-bisphosphate carboxylase/oxygenase and consequences for kinetic parameters', Plant, Cell and Environment, 36(9), pp. 1586–1596. doi: 10.1111/pce.12066.

Tcherkez, G. G. B., Farquhar, G. D. and Andrews, T. J. (2006) 'Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized', Proceedings of the National Academy of Sciences of the United States of America, 103(19), pp. 7246–7251. doi: 10.1073/pnas.0600605103.

Tung Ho, L. S. and Ané, C. (2014) 'A linear-time algorithm for gaussian and non-gaussian trait evolution models', Systematic Biology, 63(3), pp. 397–408. doi: 10.1093/sysbio/syu005.

Villar, K. Del et al. (1997) 'Amino acid substitutions that convert the protein substrate specificity of farnesyltransferase to that of geranylgeranyltransferase type I', Journal of Biological Chemistry. doi: 10.1074/jbc.272.1.680.

Wagner, P. J. (1995) 'Testing Evolutionary Constraint Hypotheses with Early Paleozoic Gastropods', Paleobiology. doi: 10.1017/S0094837300013294.

Whitney, S. M., Houtz, R. L. and Alonso, H. (2011) 'Advancing our understanding and capacity to engineer nature's CO2-sequestering enzyme, Rubisco', Plant Physiology, 155(1), pp. 27–35. doi: 10.1104/pp.110.164814.

Wilson, R. H. et al. (2018) 'An improved Escherichia coli screen for Rubisco identifies a protein-protein interface that can enhance CO2-fixation kinetics', Journal of Biological Chemistry. doi: 10.1074/jbc.M117.810861.

Wilson, R. H., Alonso, H. and Whitney, S. M. (2016) 'Evolving Methanococcoides burtonii archaeal Rubisco for improved photosynthesis and plant growth', Scientific Reports. doi: 10.1038/srep22284.

Wolfe, K. H., Li, W. H. and Sharp, P. M. (1987) 'Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs.', Proceedings of the National Academy of Sciences of the United States of America. doi: 10.1073/pnas.84.24.9054.

Yu, G. et al. (2017) 'ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data', Methods in Ecology and Evolution. doi: 10.1111/2041-210X.12628.

Zhou, Y. and Whitney, S. (2019) ‘Directed evolution of an improved Rubisco; in vitro analyses to decipher fact from fiction’, International Journal of Molecular Sciences, 20(20). doi: 10.3390/ijms20205019.

Chapter 3. Rubisco is evolving for improved catalytic efficiency and CO_2 assimilation in plants

Chapter 3, Contributors:

Bouvier, J.W., Emms, D.M. and Kelly, S.

Chapter 3, Author contributions:

J.W.B., D.M.E., and S.K. designed the research, J.W.B. and S.K. conceived the study, analysed the data, and wrote the manuscript, J.W.B. performed the research.

This thesis chapter has been published at the following location:

Bouvier, J.W., Emms, D.M. and Kelly, S. 2024. Proceedings of the National Academy of Sciences, 121(11), Article No e2321050121.

All supplemental data files are provided in the online version of this manuscript.

3.1 Abstract

Rubisco is the primary entry point for carbon into the biosphere. However, rubisco is widely regarded as inefficient leading many to question whether the enzyme can adapt to become a better catalyst. Through a phylogenetic investigation of the molecular and kinetic evolution of Form I rubisco we uncover the evolutionary trajectory of rubisco kinetic evolution in angiosperms. We show that *rbcL* is among the 1% of slowest evolving genes and enzymes on Earth, accumulating one nucleotide substitution every 0.9 million years and one amino acid mutation every 7.2 million years. Despite this, rubisco catalysis has been continually evolving toward improved CO_2/O_2 specificity, carboxylase turnover, and carboxylation efficiency. Consistent with this kinetic adaptation, increased rubisco

evolution has led to a concomitant improvement in leaf-level CO_2 assimilation. Thus, rubisco has been slowly but continually evolving toward improved catalytic efficiency and CO_2 assimilation in plants.

3.2 Introduction

Ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) converts atmospheric CO_2 into sugars that fuel the majority of life on Earth. The enzyme evolved ~3 billion years ago when the atmosphere contained high levels of CO_2 (≥10,000% present atmospheric levels) and comparatively little O_2 (≤0.1% present atmospheric levels) (Figure 3.1) (Kasting, 1993; Canfield, 2005; Kaufman *et al.*, 2007; Nisbet *et al.*, 2007; Nisbet and Nisbet, 2008; Shih *et al.*, 2016; Ślesak, Ślesak and Kruk, 2017). Since emergence, the enzyme has helped guide the atmosphere on a trajectory of increasing O_2 and declining CO_2 (Holland, 2002; Canfield, 2005) such that current concentrations of CO_2 (0.04%) and O_2 (20.95%) are inverted compared to when the enzyme first evolved (Figure 3.1).

Although all extant rubisco are descended from a single ancestral rubisco-like protein (Ashida, Danchin and Yokota, 2005; Ashida *et al.*, 2008; Erb and Zarzycki, 2018), the enzyme is found in a variety of compositional forms across the tree of life (Figure 3.1) (Tabita, Satagopan, *et al.*, 2008; Banda *et al.*, 2020). The simplest manifestations are the Form II and Form III variants found in protists, archaea, and some bacteria which are composed of a dimer, or dimers, of the ~50 kDa rubisco large (RbcL) subunit (Andersson and Backlund, 2008; Tabita, Hanson, *et al.*, 2008; Tabita, Satagopan, *et al.*, 2008; Whitney, Houtz and Alonso, 2011). In contrast, Form I rubisco is a hexadecamer comprised of four RbcL dimers organised in an antiparallel core capped at either end by the ~15 kDa rubisco small subunit (RbcS) (Schneider, Lindqvist and Brändén, 1992; Tabita, Satagopan, *et al.*, 2008). Of these three Forms, only Form I and II have been recruited for oxygenic photosynthesis (Tabita, Hanson, *et al.*, 2008), with Form I being responsible for the vast majority of global CO_2 assimilation (Field *et al.*, 1998; Tabita, Hanson, *et al.*, 2008).

Within Form I rubisco the active site is located in RbcL (Andrews, 1988; Lee and Tabita, 1990; Whitney, Houtz and Alonso, 2011). As a result, interspecific differences in Form I kinetics are

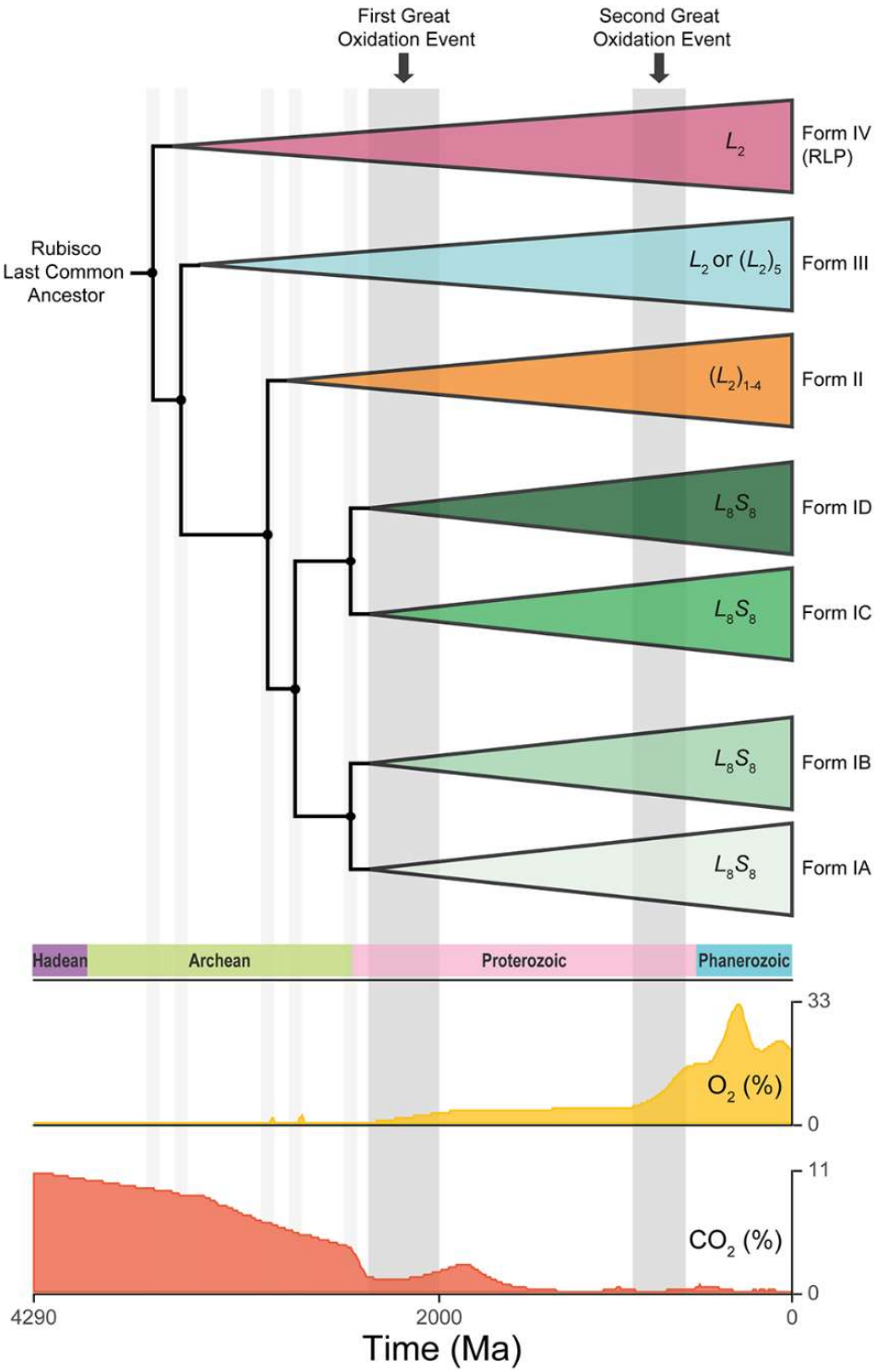


Figure 3.1. The evolutionary history of rubisco in the context of atmospheric CO_2 (%) and O_2 (%) following divergence from the ancestral rubisco-like protein (RLP). Important branch points in the phylogeny at which rubisco diverged into different evolutionary lineages are indicated by grey vertical bars. To provide additional context, the time-period at which the First and Second Great Oxidation events occurred along this evolutionary trajectory are also labelled and referenced as grey vertical bars. Graphics of atmospheric CO_2 and O_2 levels were adapted from the *TimeTree* resource (http://www.timetree.org; (Kumar *et al.*, 2017)).

primarily attributable to sequence variation in RbcL (Parry *et al.*, 2003; Kapralov and Filatov, 2007; Christin *et al.*, 2008; Whitney *et al.*, 2011; Chatterjee and Anindya, 2011; Kapralov *et al.*, 2011; Galmés, Andralojc, *et al.*, 2014; Galmés, Kapralov, *et al.*, 2014; Bouvier *et al.*, 2021; Lin *et al.*, 2021; Bouvier and Kelly, 2023; Zhou *et al.*, 2023). Despite not playing a direct role in catalysis, RbcS influences the function of rubisco (Spreitzer, 2003; Orr and Parry, 2020), and its incorporation in the holoenzyme enables its higher kinetic efficiency (Schulz *et al.*, 2022). Specifically, RbcS enhances the stability and assembly of the holoenzyme complex (Andrews, 1988; McFadden and Small, 1988; Fitchen *et al.*, 1990; Lee, Berka and Tabita, 1991; Genkov and Spreitzer, 2009; Grabsztunowicz *et al.*, 2015; Joshi *et al.*, 2015), improves the efficiency of CO_2 binding (Schneider *et al.*, 1990), and is thought to act as a reservoir for CO_2 accumulation (Van Lun *et al.*, 2014). Accordingly, rubisco function is altered when RbcS is mutated (Voordouw *et al.*, 1987; Read and Tabita, 1992b; Kostov, Small and McFadden, 1997), or when chimeric holoenzymes are created *in vivo* (Getzoff *et al.*, 1998; Kanevski *et al.*, 1999; Ishikawa *et al.*, 2011; Matsumura *et al.*, 2020; Sakoda *et al.*, 2021) and *in vitro* (Andrews and Lorimer, 1985; van der Vies, Bradley and Gatenby, 1986; Lee, Read and Tabita, 1991; Read and Tabita, 1992a; Horken and Tabita, 1999; Wang *et al.*, 2001). Moreover, there is increasing recognition of the importance of both environment (Cavanagh, Slattery and Kubien, 2022) and organ-specific (Morita *et al.*, 2014; Martin-Avila *et al.*, 2020) differences in plant RbcS isoform expression on holoenzyme catalysis. However, even though RbcS influences holoenzyme function, sequence variation in RbcL remains the primary determinant of variation in kinetics (Parry *et al.*, 2003; Kapralov and Filatov, 2007; Christin *et al.*, 2008; Whitney *et al.*, 2011; Chatterjee and Anindya, 2011; Kapralov *et al.*, 2011; Galmés, Andralojc, *et al.*, 2014; Galmés, Kapralov, *et al.*, 2014; Bouvier *et al.*, 2021; Lin *et al.*, 2021; Bouvier and Kelly, 2023; Zhou *et al.*, 2023).

Although there is kinetic variability between rubisco orthologs, the enzyme is considered to be an inefficient catalyst. For example, the maximum substrate-saturated turnover rate of Form I rubisco ($<12\ s^{-1}$) (Badger *et al.*, 1998) is slower than average (Bar-Even *et al.*, 2011). In addition, rubisco catalyses a reaction with O_2 (Bowes, Ogren and Hageman, 1971; Chollet, 1977; Sharkey, 2020) that is competitive with CO_2 and results in the loss of fixed carbon via photorespiration (Eckardt, 2005; Sharwood, 2017; Busch, 2020). As a consequence, rubisco appears poorly suited to the current O_2-

rich/CO_2-poor atmosphere (Figure 3.1). Moreover, it appears that instead of improving enzyme function, multiple lineages have evolved alternative strategies to overcome rubisco's shortcomings. For example, higher rates of CO_2 assimilation are often achieved either by synthesising large quantities of rubisco (~50% of soluble protein in plants (Carmo-Silva *et al.*, 2015) and some microbes (Sarles and Tabita, 1983; Tabita, 1999)), or by operating CO_2-concentrating mechanisms (Meyer and Griffiths, 2013; Flamholz and Shih, 2020; Schlüter *et al.*, 2023). As a result, many have questioned whether the enzyme is already perfectly adapted, and whether further kinetic improvements are possible (Ogren, 1984; Parry *et al.*, 2007, 2013; Whitney, Houtz and Alonso, 2011; Sharwood, Ghannoum and Whitney, 2016; Orr *et al.*, 2017; Sharwood, 2017; Sharkey, 2020). Obtaining answers to these questions would shed light on the "rubisco paradox" – helping to explain why this enzyme of such paramount importance appears poorly adapted for its role.

The initial hypothesis that attempted to explain the above rubisco paradox proposed that rubisco is constrained by catalytic trade-offs that limit the enzyme's adaptation. This theory was pioneered by two studies (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010) which found antagonistic correlations between rubisco kinetic traits and proposed that these trade-offs were caused by constraints on its catalytic mechanism. However, recent evidence has questioned this hypothesis as the sole mechanism to explain the rubisco paradox. Specifically, analysis of larger species sets have revealed that kinetic trait correlations are not strong (Flamholz *et al.*, 2019; Iñiguez *et al.*, 2020; Cummins, 2021). In addition, phylogenetic signal in rubisco kinetics causes kinetic trait correlations to be overestimated unless phylogenetic comparative methods are employed (Bouvier *et al.*, 2021; Bouvier and Kelly, 2023). Thus, when larger datasets are analysed with phylogenetic methods, the strength of catalytic trade-offs are substantially reduced (Bouvier *et al.*, 2021; Bouvier and Kelly, 2023). Instead, phylogenetic constraints have had a larger impact on limiting enzyme adaptation compared to catalytic trade-offs (Bouvier *et al.*, 2021; Bouvier and Kelly, 2023). These recent findings motivate a revaluation of the rubisco paradox, and an interrogation of whether rubisco is evolving for improved catalysis and CO_2 assimilation in plants.

Here, we address these issues through a phylogenetic interrogation of the molecular and kinetic evolution of the Form I holoenzyme. We reveal that RbcL has evolved at a slower rate than >98% of

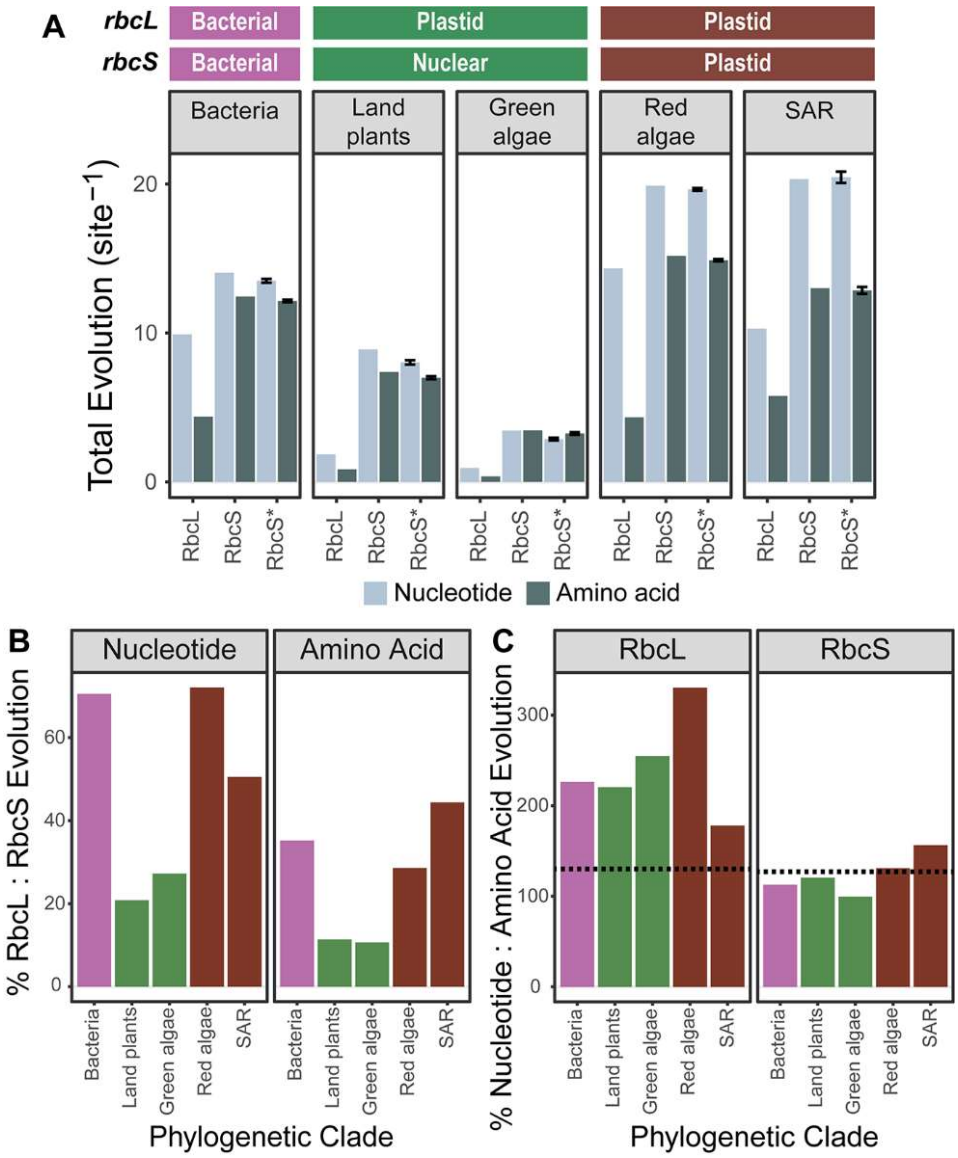


Figure 3.2. The extent of molecular evolution in rubisco during the radiation of each taxonomic group. **A)** Bar plot depicting the total amount of molecular evolution (substitutions per sequence site) in the nucleotide and protein sequences of Form I rubisco across taxonomic groups. RbcL: the extent of sequence evolution in the rubisco large subunit. RbcS: the extent of sequence evolution in the rubisco small subunit. RbcS*: the extent of sequence evolution in the rubisco small subunit using 1,000 bootstrapped stratified sampling of *rbcS*/RbcS per species (see Methods). The genome in which *rbcL* and *rbcS* genes reside within each group is indicated above the plot (bacterial, plastid, nuclear). Error bars represent ± 1 S.D of the mean. **B)** Bar plot depicting the percentage ratio (%) of nucleotide and amino acid evolution between rubisco subunits (*rbcL* to *rbcS* and RbcL to RbcS, respectively) in each taxonomic group. The colour of each bar is determined by the genome in which the *rbcL* and *rbcS* gene resides, following the colour scale in (A). **C)** Bar plot depicting the percentage ratio (%) of nucleotide to amino acid evolution in each rubisco subunit (*rbcL* to RbcL and *rbcS* to RbcS, respectively) in each taxonomic group. The colour of each bar is the same as described in (B). Dashed lines indicate the expected ratio given an *rbcL* or *rbcS* sequence evolving in the absence of selection.

all other gene/protein sequences across the tree of life. Through simultaneous evaluation of molecular and kinetic evolution of rubisco during the radiation of C_3 angiosperms, we reveal that the enzyme has been continually evolving toward improved CO_2/O_2 specificity, carboxylase turnover rate, and carboxylation efficiency. Furthermore, we demonstrate that enhanced rubisco evolution is associated with increased rates of CO_2 assimilation and higher photosynthetic nitrogen-use efficiencies. Thus, rubisco is not perfectly adapted, but is slowly evolving toward improved catalytic efficiency and CO_2 assimilation.

3.3 Results

3.3.1 RbcL has evolved slower than RbcS and has experienced stronger purifying selection

Sequences encoding Form I rubisco were obtained from the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/). This dataset was filtered to retain sequences for a given species only if a full-length sequence for both *rbcL* and *rbcS* were present. Although *rbcL* exists as a single copy gene in all species, many species harbour multiple *rbcS* genes in their genomes. Thus, for each species a single *rbcL* sequence and all available *rbcS* sequences were taken forward. In total, this resulted in a set of 488 *rbcL*/RbcL and 1140 *rbcS*/RbcS sequences across 488 species (Appendix C [and ***Supplemental File 1 online***], Figure S1 and table S1).

In order to compare the rate at which the two rubisco subunits have evolved, species were partitioned into distinct taxonomic groups comprising the red algae (*Rhodophyta*; *n* = 201), the SAR supergroup (*Stramenopiles*, *Alveolates*, and *Rhizaria*; *n* = 129), the bacteria (*Bacteria*; *n* = 78), the land plants (*Streptophyta*; *n* = 68) and the green algae (*Chlorophyta*; *n* = 12) (Appendix C [and ***Supplemental File 1 online***], Figure S1 and table S1). Hereinafter, the total amount of molecular evolution of the nucleotide sequences (*rbcL* and *rbcS*) and the total amount of molecular evolution of the protein sequences (RbcL and RbcS) in a taxonomic group is referred to as “the extent of nucleotide evolution” and “the extent of protein evolution”, respectively. The term “the extent of molecular evolution” jointly refers to both.

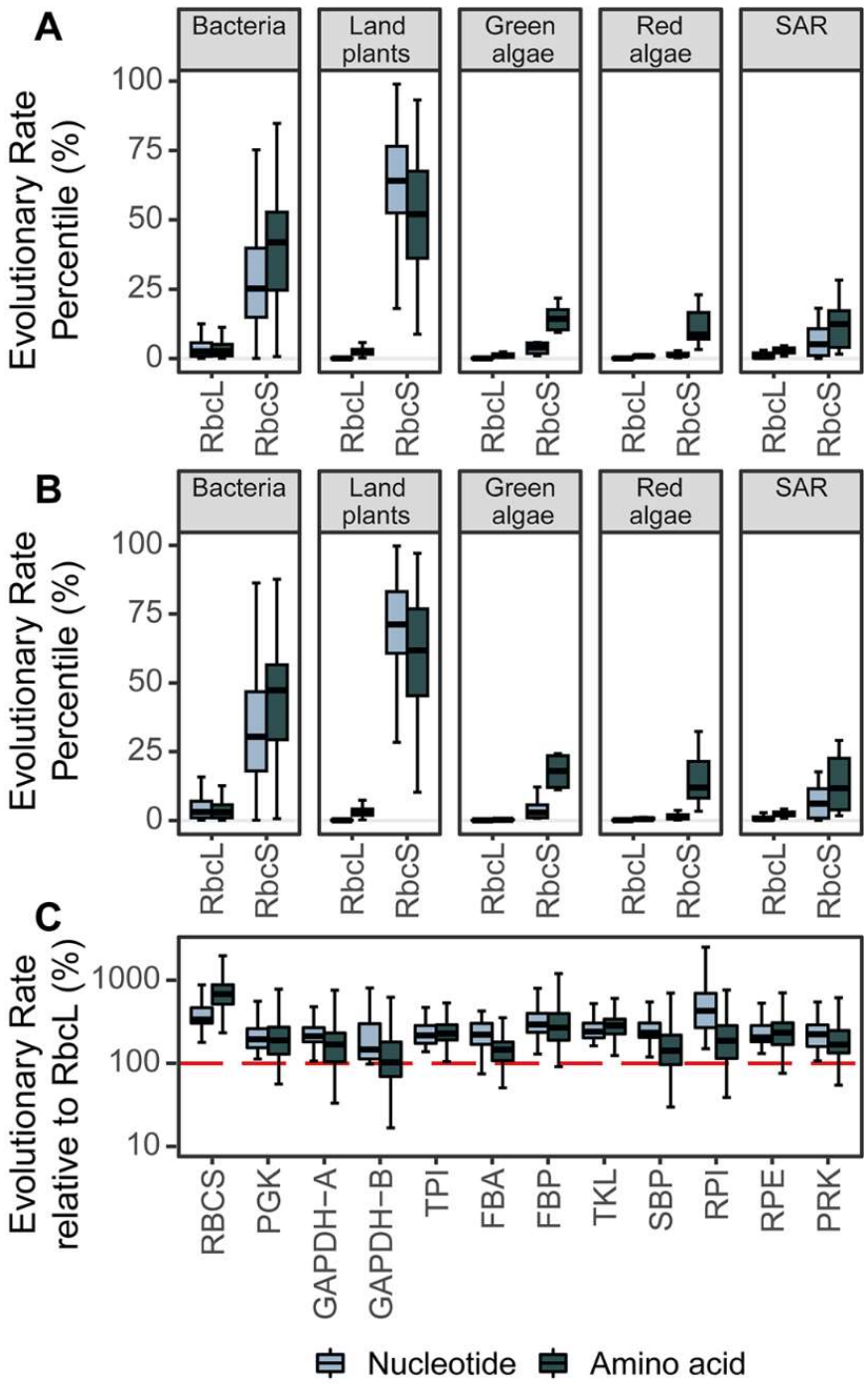


Figure 3.3. The extent of molecular evolution in rubisco in context other genes. **A)** Boxplot of the extent of molecular evolution (substitutions per sequence site) in the nucleotide and protein sequences of the rubisco large (*rbcL*/RbcL) and small (*rbcS*/RbcS) subunit expressed as a percentile (%) of that measured across all other genes and proteins, respectively. See also Appendix C [and ***Supplemental File 1 online***], table S2. **B)** As in (A) but calculating the percentile (%) extent of rubisco molecular evolution (substitutions per sequence site) relative to only the subset of genes and proteins in each species which encode enzymes. See also Appendix C [and ***Supplemental File 1 online***], table S3. **C)** Boxplot of the total amount of molecular evolution (substitutions per sequence site) in the nucleotide and protein sequences of each Calvin-Bensen-Bassham cycle enzyme expressed as a percentage (%) of that measured in *rbcL*/RbcL (100%; red horizontal line) across

land plants. Phosphoglycerate kinase: PGK. Glyceraldehyde-3-phosphate dehydrogenase A/B subunit: GAPDH-A/GAPDH-B. Triose phosphate isomerase: TPI. Fructose-bisphosphate aldolase: FBA. Fructose-1,6-bisphosphatase: FBP. Transketolase: TKL. Sedoheptulose-bisphosphatase: SBP. Ribose 5-phosphate isomerase: RPI. Ribulose-p-3-epimerase: RPE. Phosphoribulokinase: PRK. See also Appendix C [and ***Supplemental File 1 online***], table S4 and S5. The raw data for this figure can be found in ***Supplemental File 5 online***. **[*End of Figure Legend*].**

Comparison of the two rubisco subunits revealed that the extent of molecular evolution in *rbcL*/RbcL is lower than that experienced by *rbcS*/RbcS (Figure 3.2A). Specifically, the nucleotide and protein sequences of *rbcL*/RbcL have evolved at a rate equivalent to ~50% and ~25% of *rbcS*/RbcS on average across all taxa, respectively (Figure 3.2A and 3.2B). This was not an artefact of the higher gene copy number of *rbcS,* as a 1,000 bootstrapped stratified sampling recovered the same result when only a single *rbcS*/RbcS sequence was randomly sampled per species (see Methods; Figure 3.2A). Therefore, *rbcL*/RbcL has explored less nucleotide and protein sequence space than *rbcS*/RbcS in the same sets of species over the same period of time (Figure 3.2A and 3.2B). Furthermore, *rbcL* also experienced fewer amino acid changes per change in nucleotide sequence compared to *rbcS* (Figure 3.2A and 3.2C), indicating a higher degree of purifying selection. Thus, *rbcL*/RbcL has evolved more slowly and has been subject to a higher degree of functional constraint on the encoded protein sequence than *rbcS*/RbcS.

3.3.2 RbcL is one of the slowest evolving genes in the tree of life

To evaluate the rate of molecular evolution in context of all other genes in the species under consideration, the percentile rank of *rbcL*/RbcL and *rbcs*/Rbcs was evaluated for all genes in all species (see Methods). This revealed that 99.3% of all gene nucleotide sequences and 98.1% of all gene protein sequences evolved faster than *rbcL*/RbcL in the same sets of species over the same period of time (Figure 3.3A; Appendix C [and ***Supplemental File 1 online***], table S2). This held true even if *rbcL*/RbcL was only compared to the subset of genes that encode enzymes, with 99.2% of enzyme nucleotide sequences and 98.3% of enzyme protein sequences evolving faster than *rbcL*/RbcL (Figure 3.3B; Appendix C [and ***Supplemental File 1 online***], table S3). Furthermore, in land plants *rbcL*/RbcL was also the slowest evolving component of the Calvin-Benson-Bassham

cycle (Figure 3.3C; Appendix C [and ***Supplemental File 1 online***], table S4 and S5). This slow pace of evolution is not simply an artefact of being encoded in the plastid genome, as *rbcL*/RbcL was also one of the slowest evolving genes/proteins in bacteria which encode all of their genes in a single cytoplasmic genome. Thus, *rbcL*/RbcL is one of the slowest evolving genes/enzymes in all species in which it is found, irrespective of the taxonomic group or genome in which it is encoded.

In contrast to *rbcL*/RbcL, considerable variability in the extent of molecular evolution in the small subunit was observed both within and between taxonomic groups (Figure 3.3A; Appendix C [and ***Supplemental File 1 online***], table S2). Analogous results in each taxonomic group were recovered when this analysis was restricted to the subset of genes that encode enzymes (Figure 3.3B; Appendix C [and ***Supplemental File 1 online***], table S3). Moreover, in land plants *rbcS*/RbcS was the fastest evolving component of the Calvin-Benson-Bassham cycle (Figure 3.3C; Appendix C [and ***Supplemental File 1 online***], table S4 and S5). Thus, while the pace of molecular evolution in *rbcL*/RbcL is ubiquitously slow, the extent of molecular evolution of *rbcS*/RbcS is highly variable explaining the disparity in the rate of both subunits across the tree of life (Figure 3.2B; Appendix C [and ***Supplemental File 1 online***], table S6). A similar variable rate was also observed for rubisco's ancillary chaperones (Appendix C [and ***Supplemental File 1 online***]). Thus, the rate of molecular evolution of *rbcL*/RbcL is ubiquitously low, and lower than *rbcS*/RbcS or any associated chaperone.

3.3.3 Rubisco is evolving for improved kinetic efficiency in plants

Given that *rbcL* is among the slowest evolving genes on Earth, the question arises as to whether its sequence evolution is adaptive and is improving the catalysis of the enzyme. We hypothesised that if rubisco was undergoing directional selection for improved catalysis, then orthologs that have experienced the largest extent of molecular evolution would be the most efficient catalysts. To test this hypothesis, a dataset of kinetic measurements from C_3 angiosperms (Flamholz *et al.*, 2019; Bouvier *et al.*, 2021; Bouvier and Kelly, 2023) was evaluated in the context of the molecular evolution of RbcL (Figure 3.4A and 3.4B). This analysis focused on RbcL as it is the primary determinant of

Table 3.1. Rubisco kinetics in extinct and extant angiosperms. Kinetic trait values for the last common ancestor of the angiosperms were computed based on the estimated y intercept (mean ± 1 S.E.) of the linear regression analysis performed between the extent of RbcL protein evolution and each rubisco kinetic trait in Figure 3.4B. Mean values of rubisco kinetic traits and associated variation (± 1 S.E.) in extant C_3 species are shown for comparison. The raw data set used can be found in ***Supplemental File 7 online***.

Rubisco	$S_{C/O}$ (mol mol^{-1})	k_{catC} (s^{-1})	k_{catC}/K_C (s^{-1} μM^{-1})	K_C (μM)	K_C^{air} (μM)	K_O (μM)	K_C/K_O (μM μM^{-1})
Last Common Angiosperm Ancestor	81.1 ± 1.9	2.6 ± 0.3	0.16 ± 0.02	16.3 ± 2.1	24.8 ± 2.8	484.1 ± 56.4	0.034 ± 0.004
Extant Angiosperms	87.1 ± 0.5	3.4 ± 0.1	0.20 ± 0.01	17.6 ± 0.5	26.4 ± 0.7	517.2 ± 14.7	0.035 ± 0.001

kinetics (Parry *et al.*, 2003; Kapralov and Filatov, 2007; Christin *et al.*, 2008; Whitney *et al.*, 2011; Chatterjee and Anindya, 2011; Kapralov *et al.*, 2011; Galmés, Andralojc, *et al.*, 2014; Galmés, Kapralov, *et al.*, 2014; Bouvier *et al.*, 2021; Lin *et al.*, 2021; Bouvier and Kelly, 2023; Zhou *et al.*, 2023), and because sufficient sequence data for RbcS are unavailable. This revealed that the more RbcL has evolved from the most recent common ancestral sequence, the better its CO_2/O_2 specificity ($S_{C/O}$; 10.1% variance explained, $p < 0.01$), CO_2 turnover rate (k_{catC}; 4.6% variance explained, $p < 0.05$) and carboxylation efficiency (k_{catC}/K_C; 3.8% variance explained, $p < 0.05$) (Figure 3.4B). This result is not an artefact caused by potential systematic methodological biases associated with species sampling or potential uncertainties or errors in the underlying phylogenetic tree (Appendix C [and ***Supplemental File 1 online***]). Thus, rubisco has been adaptively evolving for improved $S_{C/O}$, k_{catC}, and k_{catC}/K_C during the radiation of the angiosperms.

Given that the origin of the angiosperms is estimated to have occurred 160 million years ago (Kumar *et al.*, 2017) (Figure 3.4A), it is possible to put the above kinetic change in context of both molecular sequence changes and evolutionary time since the last common ancestor at the base of this clade (table 3.1). As the large subunit acquired one nucleotide substitution every 0.9 million years and one amino acid substitution every 7.2 million years (Appendix C [and ***Supplemental File 1 online***],

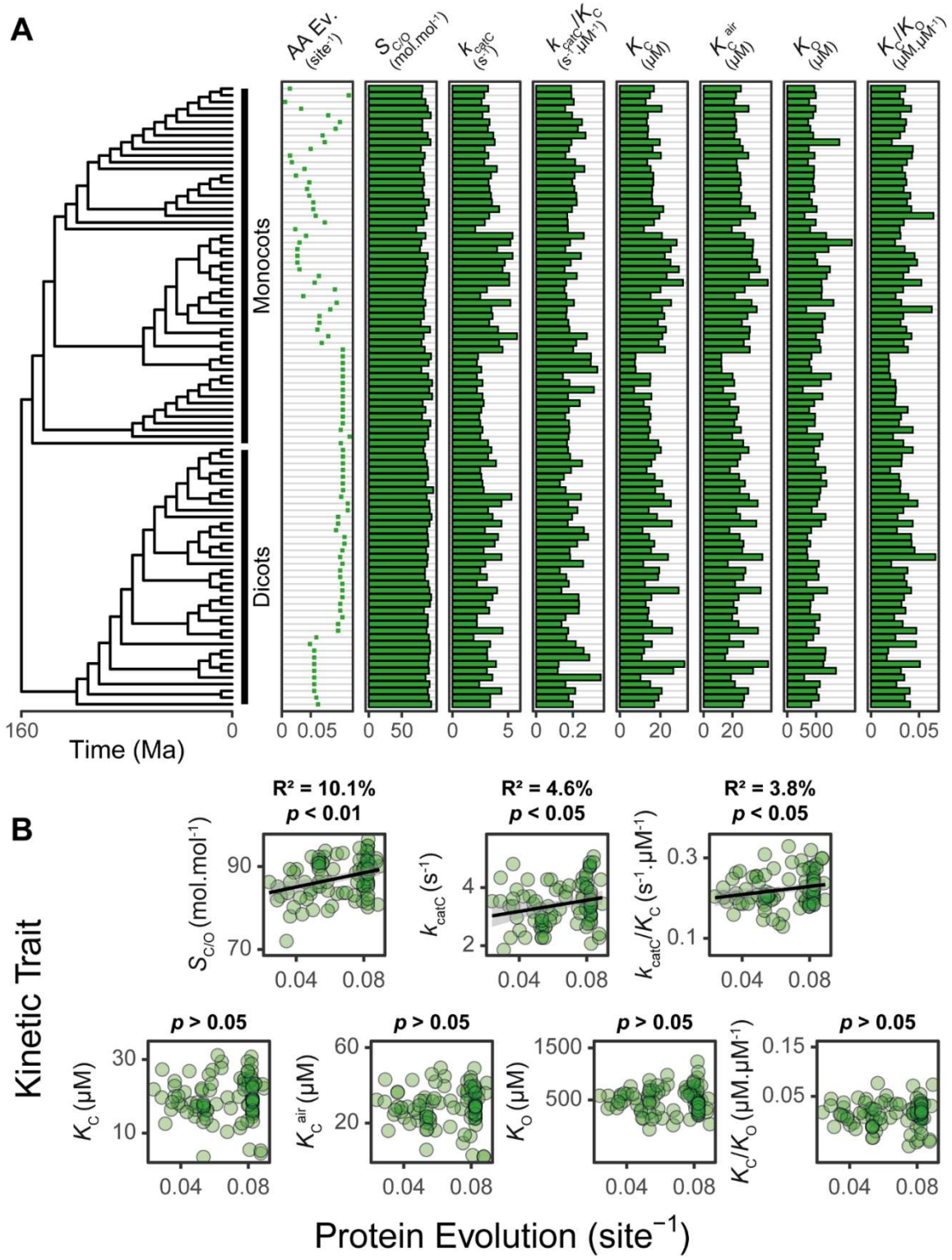


Figure 3.4. The relationship between rubisco molecular and kinetic evolution in C_3 angiosperms. **A)** The relationship between RbcL evolution and its corresponding kinetic trait values. AA Ev.: The extent of RbcL amino acid evolution that has occurred since the last common ancestor at the root of the angiosperm phylogeny. $S_{C/O}$: specificity. k_{catC}: carboxylase turnover per site. k_{catC}/K_C: carboxylation efficiency. K_C: the Michaelis constant for CO_2. K_C^{air}: the inferred Michaelis constant for CO_2 in 20.95% O_2. K_O: the Michaelis constant for O_2. K_C/K_O: the ratio of the Michaelis constant for CO_2 compared to O_2. **B)** The relationship between the extent of RbcL protein evolution (substitutions

per sequence site) and each rubisco kinetic trait in (A) as assessed using least squares regression models. The raw data can be found in ***Supplemental File 7 online***. **[*End of Figure Legend*].**

Figure S2), each amino acid substitution resulted in an average increase in $S_{C/O}$ by 2.7×10^{-1} mol.mol^{-1}, in k_{catC} by 3.6×10^{-2} s^{-1}, and in k_{catC}/K_C by 1.8×10^{-3} s^{-1} μM^{-1}. This is equivalent to a relative improvement of 0.3% ($S_{C/O}$), 1.4% (k_{catC}), and 1.1% (k_{catC}/K_C) per amino acid substitution, and a relative improvement of 0.05% ($S_{C/O}$), 0.2% (k_{catC}), and 0.2% (k_{catC}/K_C) per million years. Thus, there has been improvement in rubisco kinetics during the radiation of the angiosperms at a rate that is dependent on the extent of its molecular sequence change.

3.3.4 Rubisco is evolving for improved leaf-level CO_2 assimilation

Given that rubisco is evolving to become a better catalyst, we hypothesised that this adaptation would also drive adaptation in the rate of leaf-level CO_2 assimilation. To test this hypothesis we analysed a large dataset of photosynthetic measurements from C_3 angiosperms (Gago *et al.*, 2019) in the context of the extent of their RbcL evolution (Figure 3.5A – 3.5C). This revealed that the rate of leaf-level CO_2 assimilation was also dependent on the extent of molecular sequence change in rubisco, such that C_3 angiosperms with more evolved rubisco also presented higher rates of CO_2 assimilation (A_{mass}; 19.2% variance explained, $p < 0.001$) (Figure 3.5B). This is not a consequence of increased nitrogen investment in the leaf, as the association between rubisco evolution and increased CO_2 assimilation is strengthened when measurements are controlled for leaf nitrogen content ($PNUE_{mass}$, 22.1% variance explained, $p < 0.001$) (Figure 3.5B). Analogous results were obtained when measurements of CO_2 assimilation were evaluated on a leaf area basis (Figure 3.5C). This result is most parsimoniously explained by directional selection towards enhanced leaf-level CO_2 assimilation driven by the kinetic adaptation described above. Thus, the adaptive evolution of rubisco during the radiation of the angiosperms has resulted in the improvement in leaf-level CO_2 assimilation.

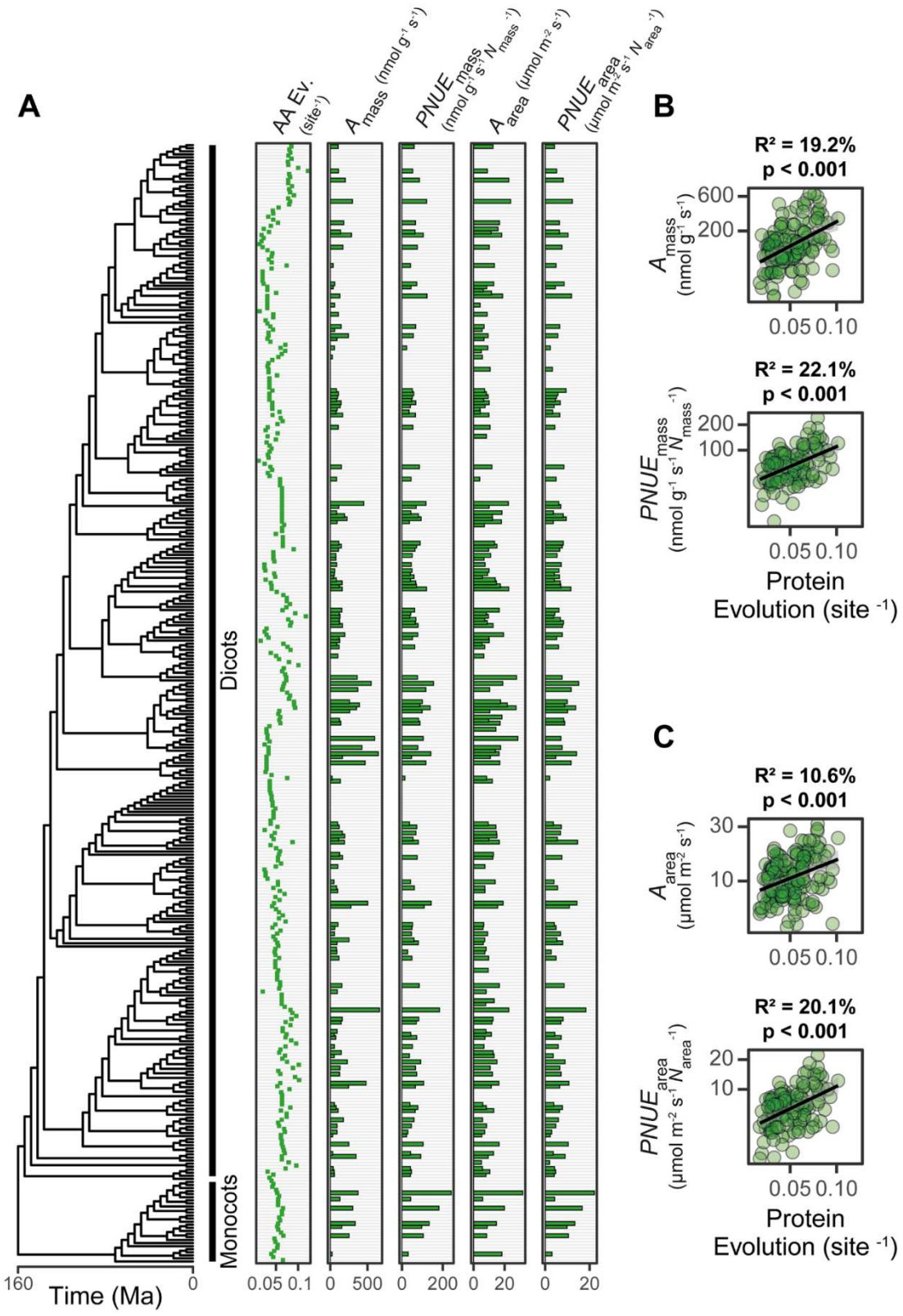


Figure 3.5. The relationship between rubisco molecular evolution and CO_2 assimilation in C_3 angiosperms. **A)** The relationship between the extent of RbcL evolution and leaf level CO_2 assimilation. AA Ev.: The extent of RbcL amino acid evolution that has occurred since the most recent common ancestor at the root of the angiosperm phylogeny. A_{mass}: Photosynthetic rate per unit leaf mass. $PNUE_{mass}$: Photosynthetic nitrogen use efficiency rate, calculated as photosynthetic rate per unit leaf mass per unit leaf nitrogen content (N_{mass}; % N). A_{area}: Photosynthetic rate per unit leaf

area. $PNUE_{area}$: Photosynthetic nitrogen use efficiency rate, calculated as photosynthetic rate per unit leaf area per unit leaf area nitrogen content (N_{area}; g m^{-2} N). **B)** The relationship between the extent of RbcL protein evolution (substitutions per sequence site) and each photosynthetic trait in (A) evaluated on a mass-basis (A_{mass}, $PNUE_{mass}$) as assessed using least squares regression models. **C)** As in (B) but for each photosynthetic trait evaluated on an area-basis (A_{area}, $PNUE_{area}$). The raw data can be found in ***Supplemental File 9 online***. **[*End of Figure Legend*].**

3.4 Discussion

Rubisco is the primary entry point for carbon into the biosphere, responsible for fixing 250 billion tons of CO_2 annually (Field *et al.*, 1998). Despite this immense throughput, the enzyme is a surprisingly inefficient catalyst with a modest carboxylase turnover rate of <12 s^{-1} (Badger *et al.*, 1998) and a competing oxygenase activity that results in the loss of fixed carbon (Bowes, Ogren and Hageman, 1971; Ogren and Bowes, 1971; Sharkey, 2020). This discord presents an evolutionary paradox that has attracted significant attention (Ogren, 1984; Parry *et al.*, 2007, 2013; Whitney, Houtz and Alonso, 2011; Sharwood, Ghannoum and Whitney, 2016; Orr *et al.*, 2017; Sharwood, 2017; Sharkey, 2020; Bouvier *et al.*, 2021; Bouvier and Kelly, 2023). Here we demonstrate *rbcL* is one of the slowest evolving genes on Earth. Despite this, we demonstrate that rubisco has been evolving for higher CO_2/O_2 specificity ($S_{C/O}$), faster carboxylase turnover rates (k_{catC}), and improved carboxylation efficiencies (k_{catC}/K_C) in angiosperms. Moreover, we demonstrate that plants with more evolved rubisco exhibit higher leaf-level CO_2 assimilation and enhanced photosynthetic nitrogen-use efficiencies. Thus, rubisco has been continually evolving towards improved catalytic efficiency and CO_2 assimilation during the radiation of the angiosperms.

A slow rate of molecular evolution in *rbcL* has long been assumed and has underpinned the use of this gene for systematics and phylogenetics (Gielly and Taberlet, 1994; APG, 1998, 2016). However, to our knowledge there has been no contextualised measurement of the rate of *rbcL* evolution across the tree of life. The analysis presented here addresses this gap by revealing that *rbcL*/RbcL has experienced a lower extent of molecular evolution than 99% of all gene nucleotide sequences and 98% of all gene protein sequences across all taxa in which it is encoded. Whilst *rbcL*/RbcL evolves

at a moderate pace relative to the handful of other genes encoded in the angiosperm chloroplast genome (Appendix C, Figure S3A and S3B), it is interesting to note that the slow rate of *rbcL* evolution is not due to the presence of *rbcL* in the chloroplast genome. This is because *rbcL* is also one of the slowest evolving sequences in bacteria which lack organellar genomes. Thus, RbcL is universally one of the slowest evolving sequences on Earth, irrespective of the taxon or genome in which it resides.

Although dissecting the factors which constrain the rate of *rbcL* evolution is beyond the scope of the current study, the slow pace of *rbcL* molecular evolution is most likely a consequence of several synergistic factors (Pál, Papp and Lercher, 2006) including constraints imposed by expression (Drummond *et al.*, 2005; Wall *et al.*, 2005; Gout, Kahn and Duret, 2010; Seward and Kelly, 2018; Robbins and Kelly, 2023), selection to preserve protein function (Tourasse and Li, 2000; Bloom *et al.*, 2006; Lobkovsky, Wolf and Koonin, 2010; Soskine and Tawfik, 2010; Echave, 2019), and the requirements for protein-protein interactions *in vivo* (Fraser, Wall and Hirsh, 2003; Fraser and Hirsh, 2004; Saeed and Deane, 2006; Tóth-Petróczy and Tawfik, 2011). These factors would be particularly pertinent for rubisco given that it is the most abundant enzyme in organisms in which it is found (Tabita, 1999; Carmo-Silva *et al.*, 2015), it is subject to catalytic trade-offs (Tcherkez, Farquhar and Andrews, 2006; Flamholz *et al.*, 2019; Bouvier *et al.*, 2021; Bouvier and Kelly, 2023) and molecular activity-stability trade-offs (Studer *et al.*, 2014; Duraõ *et al.*, 2015; Cummins, Kannappan and Gready, 2018, 2019), and given that it relies on multiple interacting partners and chaperones for folding, assembly and metabolic regulation (Carmo-Silva *et al.*, 2015; Aigner *et al.*, 2017). Thus, a perfect storm of features exist which could limit the molecular evolution of *rbcL* and thereby cause it to be one of the slowest evolving genes on Earth. Further work to elucidate the exact contribution of each of these biological determinants on rubisco's rate of molecular evolution is warranted, building upon the work here and previous investigations (Whitney *et al.*, 2015; Bouvier *et al.*, 2021; Bouvier and Kelly, 2023).

Our integrated analysis of rubisco evolution revealed a continual improvement in $S_{C/O}$, k_{catC} and k_{catC}/K_C during the radiation of C_3 angiosperms. Thus, although rubisco is slowly-evolving, sequence changes have enhanced the catalytic properties of the enzyme. In the context of the C_3 leaf, such

directional selection towards improved $S_{C/O}$ is consistent with adaptation to maintain adequate carbon assimilation in response to declining CO_2 and increasing O_2 (Figure 3.1). This evolutionary strategy has been proposed previously (Tcherkez, Farquhar and Andrews, 2006), and is suggested to apply broadly across photoautotrophs lacking a CO_2 concentrating mechanism (Rickaby and Eason Hubbard, 2019; Iñiguez *et al.*, 2020). In addition to adaptation for higher $S_{C/O}$, we also discover simultaneous improvement in k_{catC} and k_{catC}/K_C without antagonism in any other kinetic trait. These results are also consistent with the inferior $S_{C/O}$ and k_{catC}/K_C reported for extinct rubisco resurrected at the dawn of the Form IB (Shih *et al.*, 2016) and Form 1 (Schulz *et al.*, 2022) lineages. It is noteworthy that on first appearances, all of these studies seem to contradict an analysis within the *Solanaceae* in which resurrected ancestral rubisco variants exhibited superior k_{catC} and k_{catC}/K_C values (Lin *et al.*, 2022). However, in this instance the kinetic differences were proposed to be driven by sequence changes in RbcS (Lin *et al.*, 2022), and therefore do not contradict the analysis of RbcL presented here or in other studies (Shih *et al.*, 2016; Schulz *et al.*, 2022). Thus, sequence change in RbcL during the radiation of angiosperms has driven the continual improvement of the enzyme in the presence of a declining atmospheric CO_2:O_2 concentration.

The 'FvCB model' of photosynthesis (Farquhar, von Caemmerer and Berry, 1980), as well as a suite of other experimental (Bassham and Krause, 1969; Dietz and Heber, 1984; Quick *et al.*, 1991; Stitt *et al.*, 1991; Hudson *et al.*, 1992; Fichtner *et al.*, 1993; Furbank *et al.*, 1996; Maheshwari *et al.*, 2021) and computational (Poolman, Fell and Thomas, 2000; Zhu, De Sturler and Long, 2007) studies all demonstrate that rubisco is a major rate-limiting factor for CO_2 assimilation under ambient steady-state conditions. The findings presented here link these mechanistic studies with evolutionary biology, and reveal that rubisco has experienced directional selection to improve kinetic efficiency and CO_2 assimilation. It is worth noting that the strength of the relationship observed between rubisco evolution and leaf level CO_2 assimilation is greater in magnitude than that between evolution and enzyme kinetics. This phenomenon can be explained by the fact that rubisco is not only evolving for improved kinetic efficiency, but is also simultaneously undergoing optimisation for multiple other properties such as stability, activatability, and chaperone affinity. This multi-objective optimisation means that not all sequence changes are expected to alter kinetics. Instead, these sequence

changes influence a much larger set of properties that together contribute to improvement in leaf-level CO_2 assimilation. Thus, our finding that rubisco evolution has a stronger correlation with leaf-level CO_2 assimilation than with any individual kinetic trait aligns with evolutionary expectations.

Ultimately, the global trends of rubisco evolution presented here change our understanding of the rubisco paradox. Rubisco is not locked in evolutionary stasis, but is instead slowly evolving toward improved CO_2 assimilation. These discoveries have significant implications for our understanding of the past, present, and future potential of rubisco in natural and engineered contexts. Crucially, they also serve as a foundation for future work to further unravel the intricacies of rubisco evolution. For instance, an important focus of subsequent research would be to elucidate the adaptative trajectory of rubisco across different temporal, spatial and phylogenetic scales. The present findings also motivate efforts to resolve the contribution of the small subunit and accessory chaperones to rubisco adaptation, given that these genes evolve faster than the large subunit and thus may have played an important role in the evolutionary dynamics of the enzyme over geological time scales.

3.5 Materials and Methods

3.5.1 Rubisco sequence data

All publicly available coding sequences of the rubisco large (*rbcL*) and small (*rbcS*) subunit genes in the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/) as of July 2020 were downloaded (*rbcL* *n* = 239,492; RbcS *n* = 2,061). Manual inspection of nucleotide and translated protein sequences was performed to remove any duplicate, partial, chimeric, or erroneously annotated sequences. In addition, this dataset was further restricted to include only those species which possess a Form I rubisco and for which both a full-length *rbcL* and *rbcS* gene sequence could be obtained. Given that *rbcL* exists as a single copy gene in all species, only one *rbcL* sequence per species was retained for downstream analysis. In contrast, all possible full-length *rbcS* sequences were taken forward to account for the fact that *rbcS* is multicopy in some genomes.

Translated *rbcL* and *rbcS* protein sequences were aligned using the MAFFT L-INS-i algorithm (Katoh and Standley, 2013). The corresponding codon alignments of the nucleotide sequences were generated by threading the nucleotide sequences through the aligned protein sequences that they encode using PAL2NAL software (Suyama, Torrents and Bork, 2006). Multiple sequence alignments were trimmed to remove non-aligned codon or residue positions such that only ungapped columns remained. During this process, the putative transit peptide of *rbcS*/RbcS sequences in taxa in which this gene is encoded by the nuclear genome was computationally cleaved. Following these data processing steps, alignments were partitioned depending on species membership to either the bacteria (*Bacteria*; *n* = 78), land plants (*Streptophyta*; *n* = 68), green algae (*Chlorophyta*; *n* = 12), red algae (*Rhodophyta*; *n* = 201) or the SAR supergroup (*Stramenopiles*, *Alveolates*, and *Rhizaria;* *n* = 129) by use of the NCBI taxonomy browser (https://www.ncbi.nlm.nih.gov/Taxonomy/TaxIdentifier/tax_identifier.cgi). Any sequences belonging to species in either the *Haptophyta*, *Cryptophyta*, *Glaucocystophyta* or *Excavata* taxonomic groups were excluded from the dataset at this point due to insufficient data availability. In total, this resulted in a combined set of 488 *rbcL*/RbcL and 1140 *rbcS*/RbcS gene and protein sequences across 488 species spanning 5 taxonomic groups (Appendix C [and ***Supplemental File 1 online***], Figure S1 and table S1). The complete set of raw *rbcL*/RbcL and *rbcS*/RbcS sequences, as well as the complete set of aligned and trimmed *rbcL*/RbcL and *rbcS*/RbcS sequences can be found in ***Supplemental File 2 online***.

3.5.2 Rubisco phylogenetic tree inference

Maximum-likelihood *rbcL*/RbcL and *rbcS*/RbcS phylogenetic gene trees were inferred across all sequences within a taxonomic group by IQ-TREE (Nguyen *et al.*, 2015) using the ultrafast bootstrapping method with 1,000 replicates and the Shimodaira–Hasegawa approximate–likelihood ratio branch test. The best fitting models of nucleotide (SYM+R8) and amino acid (LG+R5) sequence evolution were respectively determined as those which exhibit the lowest combined Bayesian information criterion rank score across the complete sets of both rubisco large and small subunit sequences (Nguyen *et al.*, 2015). Across all taxonomic groups, the models of nucleotide and amino acid sequence evolution were held constant between the gene trees for *rbcL* and *rbcS*, and RbcL

and RbcS, respectively, such that branch lengths are comparable across both subunits. The complete set of these *rbcL*/RbcL and *rbcS*/RbcS phylogenetic gene trees used as the basis of the analysis herein can be found in ***Supplemental File 3 online***. To account for potential biases in our analysis caused by some species exhibiting multiple copies of rbcS, random stratified sampling of the non-gapped *rbcS*/RbcS sequence alignments was also conducted as described in Appendix C and ***Supplemental File 1 online***.

3.5.3 Quantification of the total extent of nucleotide and protein molecular evolution in rbcL/RbcL and rbcS/RbcS

The extent of molecular evolution in both rubisco subunits was assessed across all species in a given taxonomic group as the total length (sequence substitutions per aligned sequence site) of the phylogenetic tree describing the evolutionary history of each respective gene. For this purpose, tree length was calculated as the combined sum of branch lengths leading from the root at the last common ancestor of the tree to the set of sequences at the terminal nodes. In this way, using the trees inferred across the complete cohort of *rbcL*/RbcL and *rbcS*/RbcS sequences in each taxonomic group, it was possible to capture all nucleotide and amino acid evolution which has arisen in each subunit since the most recent common ancestor of all sampled species in the group. The total tree length measure used here accounts for differences in alignment length between genes because the branch lengths in the tree are estimates of the number of substitutions per sequence site. This measure is also unaffected by species sampling, or variation in model of sequence evolution, as each of these factors was held constant in the analyses that were conducted. An identical analysis was also performed for each *rbcS*/RbcS tree generated by stratified sampling, with mean and standard errors of estimates being calculated in this case across the 1,000 unique bootstrap replicate trees.

3.5.4 Orthogroup classification and phylogenetic tree inference

Complete sets of representative gene models were acquired for a total of 32 of the bacteria species, 27 of the land plant species, 8 of the SAR species, 6 of the red algae species and 4 of the green algae species analysed in the present study, respectively (Appendix C [and ***Supplemental File 1***

online], table S1 and S7). A detailed description of the method used to download and prepare these gene models for downstream analysis is provided in Appendix C [and ***Supplemental File 1 online***].

The complete set of translated proteomes for species in each respective taxonomic group were subject to orthogroup inference using OrthoFinder V2.5.2 (Emms and Kelly, 2015, 2019) software run with default settings and with the DIAMOND ultra-sensitive mode (Buchfink, Xie and Huson, 2014; Buchfink, Reuter and Drost, 2021). Protein sequences within each orthogroup were aligned using the MAFFT L-INS-I algorithm with 1,000 cycles of iterative refinement (Katoh and Standley, 2013). The corresponding codon alignments of the nucleotide sequences were generated by threading the nucleotide sequences through the aligned protein sequences that they encode using PAL2NAL software (Suyama, Torrents and Bork, 2006). Alignments were trimmed to remove positions which contain gap characters. Sequences that were <50% of the median length of the cohort of all other sequences in the given orthogroup were excluded to avoid analysis of partial or truncated genes that could influence downstream analysis. All nucleotide and protein multiple sequence alignments which satisfied the above criteria and which possessed >50 ubiquitously aligned codon or amino acid positions were subject to bootstrapped maximum likelihood phylogenetic tree inference using IQ-TREE (Nguyen *et al.*, 2015) following the exact method and evolutionary substitution models described above. In total, this resulted in a combined set of 16,631 orthogroup phylogenies comprising 5,126,017 ortholog pairwise comparisons across 351 species pairwise comparisons for the land plant clade, 6,953 orthogroup phylogenies comprising 153,288 ortholog pairwise comparisons across 28 species pairwise comparisons for the SAR clade, 5,422 orthogroup phylogenies comprising 642,057 ortholog pairwise comparisons across 496 species pairwise comparisons for the bacteria clade, 4,269 orthogroup phylogenies comprising 31,133 ortholog pairwise comparisons across 6 species pairwise comparisons for the green algae clade and 3,966 orthogroup phylogenies comprising 54,091 ortholog pairwise comparisons across 15 species pairwise comparisons for the red algae clade, from which to base the analyses herein. A further breakdown of these metrics for each species comparison can be found in ***Supplemental File 4 online***. The set of these respective genes within each species proteome that encode enzymes was determined following the method described in Appendix C [and ***Supplemental File 1 online***].

3.5.5 Quantification of the percentile extent of rubisco molecular evolution within each taxonomic group

To evaluate the extent of molecular evolution in rubisco in the context of all other genes, only species in the rubisco sequence dataset possessing a publicly available whole-genome gene assembly were considered. Across each pairwise combination of species in a given taxonomic group which satisfied this criteria, the extent of *rbcL*/RbcL and *rbcS*/RbcS molecular evolution since the time point of species divergence was measured by computing the sum of branch lengths (sequence substitutions per aligned sequence site) separating these respective sequences in the rubisco phylogenetic trees previously inferred. Following this, the extent of molecular evolution separating all other pairs of orthologous (but not paralogous) gene and protein sequences for that given species pair was measured across all inferred orthogroup phylogenies, and the percentile rank rate of rubisco nucleotide or protein evolution was computed relative to the cohort of these measurements. To assess the extent of rubisco molecular evolution in context all other enzymes, the exact same steps were followed but only the subset of genes and proteins predicted to encode enzymes were included. In both of the above analyses, a minimum threshold of 100 measurements for orthologous genes and protein sequences was ensured per species pair. In cases where multiple percentiles are calculated for a rubisco subunit in a given species pair (due to gene duplications in the *rbcS* of some species, or due to a single species gene assembly matching multiple sub-species in the rubisco sequence dataset) the mean percentile was taken. To provide further context on the relative pace of rubisco molecular evolution, the difference in evolutionary rate between rubisco and all other photosynthetic isoforms of the Calvin-Benson-Bassham cycle enzymes (Appendix C [and ***Supplemental File 1 online***], table S8) was also determined across the land plants where this data is available, as described in Appendix C [and ***Supplemental File 1 online***]. In addition, quantification of the percentile extent of molecular evolution in rubisco chaperones (Appendix C [and ***Supplemental File 1 online***], table S9) was performed following the same method and the same dataset for rubisco above, as described in Appendix C [and ***Supplemental File 1 online***].

The full set of raw data generated from these analyses measuring the molecular evolution in rubisco and across all other genes and proteins, and all other enzyme-encoding genes and proteins in each

taxa have been deposited in figshare (https://doi.org/10.6084/m9.figshare.24994625). Individual processed datasets quantifying the relative extents of molecular evolution in rubisco and rubisco chaperones for each unique pairwise species comparison can be found in ***Supplemental File 5 online***. This processed data has also been included in a combined dataset format in ***Supplemental File 6 online***.

3.5.6 Integrated analysis of rubisco molecular and kinetic evolution

To interrogate the relationship between the molecular and kinetic evolution of extant Form I rubisco, a dataset of rubisco kinetic traits was downloaded from (Bouvier *et al.*, 2021; Bouvier and Kelly, 2023), as modified from that originally compiled by Flamholz and colleagues (Flamholz *et al.*, 2019). For the purpose of this study, only species in this dataset with a complete set of experimentally determined measurements of rubisco specificity ($S_{C/O}$) for CO_2 relative to O_2 (i.e., the overall carboxylation/oxygenation ratio of rubisco under defined concentrations of CO_2 and O_2 gases), maximum carboxylase turnover rate per active site (k_{catC}), and the respective Michaelis constant (i.e., the substrate concentration at half-saturated catalysed rate) for both CO_2 (K_C) and O_2 (K_O) substrates were selected. For each of the 137 species which satisfied this criteria (all of which were angiosperm land plants), an estimate of the Michaelis constant for CO_2 in 20.95% O_2 air (K_{Cair}) was also available (Bouvier *et al.*, 2021; Bouvier and Kelly, 2023). In addition, the ratio of the Michaelis constant for CO_2 relative to O_2 (K_C/K_O), as well as carboxylation efficiency defined as the ratio of the maximum carboxylase turnover to the Michaelis constant for CO_2 (k_{catC}/K_C), were also inferred. Measurements of the Michaelis constant for RuBP (K_{RuBP}) were not considered owing to a limited sample size (n = 19). All *Limonium* species in the dataset were also removed (Galmés *et al.*, 2005; Galmés, Andralojc, *et al.*, 2014) on the basis of the arguments presented in (Tcherkez and Farquhar, 2021). In total, this left a dataset of rubisco kinetic trait measurements for 123 angiosperms. Of these, only the subset of 93 species which perform C_3 photosynthesis were considered for the purpose of the integrated molecular and kinetic evolution analysis herein. This is because of both a limited sample size of C_3–C_4 (n = 6), C_4–like species (n = 3) and C_4 species (n = 21) in the kinetic dataset, and given that transition toward C_4 photosynthesis is associated with a change in rubisco kinetic evolution (Bouvier

et al., 2021; Bouvier and Kelly, 2023) that would confound the directional selection analysis being conducted.

Coding sequences of the *rbcL* gene were obtained from Bouvier and colleagues (Bouvier *et al.*, 2021) for each species in the kinetic dataset. In order to facilitate more accurate downstream phylogenetic tree inference across these sequences and to minimize the impact of long-branch effects (Su and Townsend, 2015), the complete set of publicly available *rbcL* coding sequences in land plants were also acquired in parallel from NCBI (https://www.ncbi.nlm.nih.gov/) using the query term "rbcL[Gene Name] AND "plants"[porgn:_txid3193]". These sequences thus obtained were subject to the exact same data processing steps to remove ambiguous, partial or chimeric sequences as performed previously for the *rbcL* sequences of species in the rubisco kinetic dataset (Bouvier *et al.*, 2021). In total, this step resulted in an additional set of 29,218 full-length *rbcL* coding sequences to aid downstream phylogenetic inference. Protein sequences were inferred from each *rbcL* coding sequences via *in silico* translation. Next, the complete set of translated RbcL sequences (including the set of sequences from angiosperms in the rubisco kinetic dataset, as well as the set of all publicly available sequences for land plants) were respectively aligned using MAFFT L-INS-I (Katoh and Standley, 2013), and a corresponding *rbcL* coding sequence alignment was generated using PAL2NAL software (Suyama, Torrents and Bork, 2006). The resulting multiple sequence alignments were trimmed to remove non-aligned residue positions and bootstrapped phylogenetic trees were inferred using IQ-TREE (Nguyen *et al.*, 2015) following the exact method described above and using the best-fit models of sequence evolution previously inferred. To facilitate downstream analysis, the *rbcL* and RbcL gene trees were subsequently modified to keep only internal and terminal branches leading to the set of species in the rubisco kinetic dataset, with pruned trees manually rooted in Dendroscope (Huson and Scornavacca, 2012).

To compute the relative extent of protein evolution which has occurred in each angiosperm in the kinetic dataset, the summed branch length (sequence substitutions per aligned sequence site) leading from the last common ancestor at the root of this clade to each respective terminal node in the RbcL phylogeny were measured. The kinetic trait values and extent of molecular evolution for all C_3 angiosperm rubisco can be found in ***Supplemental File 7 online***. The predicted kinetic trait

values at the last common ancestor at the base of the angiosperm clade were inferred from the estimated y-intercept values from these regression models, as found in table 3.1. The *rbcL*/RbcL phylogenetic gene trees used as the basis of this analysis, including the trees inferred across the full set of sequences, as well as the pruned versions of these trees containing only the subset of C_3 species in the kinetic dataset, can be found in ***Supplemental File 8 online***.

To confirm that the above analysis of rubisco molecular and kinetic evolution was not an artefact of potential methodological biases, this analysis was repeated using both a minimal subset of phylogenetically diverse species (Appendix C [and ***Supplemental File 1 online***], table S10) so as to control for incomplete species sampling, as well as using alternate RbcL phylogenetic gene trees so as to control for any phylogenetic uncertainties. These supporting set additional analyses confirmed that the results of the original analysis were valid and robust to potential methodological biases. A full description of the design, method, and results of these additional analyses is provided in detail in Appendix C [and ***Supplemental File 1 online***].

3.5.7 Integrated analysis of rubisco molecular evolution and CO_2 assimilation

To investigate the relationship between rubisco molecular evolution and whole-plant photosynthetic performance, a comprehensive meta-dataset of photosynthetic measurements from species spanning the whole land plant phylogeny was provided by Gago and colleagues (Gago *et al.*, 2019). This dataset contained measurements of light-saturated net photosynthetic rates expressed both per unit leaf mass (A_{mass}) and per unit leaf area (A_{area}), as well as measurements of total nitrogen content expressed both per unit leaf mass (N_{mass}) and per unit leaf area (N_{area}). In addition, for each unique species observation in this dataset with a corresponding measurement for both A_{mass} and N_{mass} or for both A_{area} and N_{area}, the mass-based and area-based photosynthetic nitrogen-use efficiencies were also derived using the calculations A_{mass}/N_{mass} ($PNUE_{mass}$) and A_{area}/N_{area} ($PNUE_{area}$), respectively. In cases where duplicate entries for a parameter were present across species, the mean value was taken so as to collapse the dataset to contain only a single row per species. Finally, although photosynthetic measurements were available from individuals belonging to all major land plant lineages (including the mosses, liverworts, fern allies, ferns, gymnosperms, and angiosperms), only the subset of angiosperms for which a publicly available *rbcL* sequence could be obtained were

taken forward. This is because various diffusional and biochemical factors other than rubisco are known to cause reduced photosynthetic capacities in non-angiosperm plants (Gago *et al.*, 2019) that would bias the results of the current study. For the same reasons, only the subset of C_3 angiosperms in this dataset were taken forward to avoid picking up photosynthetic effects which result from CO_2 concentrating mechanisms that act upstream of rubisco. In total, this left a photosynthetic dataset of 366 C_3 angiosperms from which to base the analyses herein. This dataset included 272 unique species measurements for N_{mass}, 137 unique species measurements for A_{mass} and 118 unique species measurements for $PNUE_{mass}$, as well as 270 unique species measurements for N_{area}, 151 unique species measurements for A_{area}, and 120 unique species measurements for $PNUE_{area}$, respectively.

To compute the relative extents of RbcL molecular evolution which has occurred in each angiosperm in the photosynthetic dataset, the exact same method was followed as described above. First, the full RbcL phylogenetic gene tree in ***Supplemental File 8 online*** that was previously inferred from the complete set of publicly available RbcL sequences in NCBI was pruned so as to contain only terminal and internal branches corresponding to angiosperms in the photosynthetic dataset. Here, in situations where duplicate sequences in the alignment resulted in multiple terminal nodes for a given species, only a single node was retained based on the sequence which is first in the alphabetical order of the gene accession numbers. As above, this reduced RbcL tree was then manually rooted in Dendroscope (Huson and Scornavacca, 2012), and the relative extent of RbcL protein evolution in each angiosperm was computed as the summed branch length (sequence substitutions per aligned sequence site) leading from the last common ancestor at the root of this clade to each respective terminal node. Finally, linear regression models were employed to assess the pairwise relationships between the variation in rubisco molecular evolution and each respective photosynthetic parameter. The resulting full integrated dataset containing photosynthetic measurements and comparable extents of RbcL molecular evolution for all 366 C_3 angiosperms can be found in ***Supplemental File 9 online***. The RbcL phylogenetic gene tree which has been pruned from that in ***Supplemental File 8 online*** to contain the subset of C_3 angiosperms in the

photosynthetic dataset used for the basis of this analysis can be found in ***Supplemental File 10 online***.

3.6 References

Aigner, H. et al. (2017) 'Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2', Science, 358(6368), pp. 1272–1278. Available at: https://doi.org/10.1126/science.aap9221.

Andersson, I. and Backlund, A. (2008) 'Structure and function of Rubisco', Plant Physiology and Biochemistry, 46(3), pp. 275–291. Available at: https://doi.org/10.1016/j.plaphy.2008.01.001.

Andrews, T.J. (1988) 'Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits.', The Journal of biological chemistry, 263(25), pp. 12213–12219. Available at: https://doi.org/10.1016/s0021-9258(18)37741-x.

Andrews, T.J. and Lorimer, G.H. (1985) 'Catalytic properties of a hybrid between cyanobacterial large subunits and higher plant small subunits of ribulosebisphosphate carboxylase-oxygenase', Journal of Biological Chemistry, 260(8), pp. 4632–4636. Available at: https://doi.org/10.1016/s0021-9258(18)89117-7.

APG (1998) An Ordinal Classification for the Families of Flowering Plants, Annals of the Missouri Botanical Garden. Available at: https://doi.org/10.2307/2992015.

APG (2016) 'An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV', Botanical Journal of the Linnean Society, 181(1), pp. 1–20. Available at: https://doi.org/10.1111/boj.12385.

Ashida, H. et al. (2008) 'RuBisCO-like proteins as the enolase enzyme in the methionine salvage pathway: Functional and evolutionary relationships between RuBisCO-like proteins and photosynthetic RuBisCO', Journal of Experimental Botany, 59(7), pp. 1543–1554. Available at: https://doi.org/10.1093/jxb/ern104.

Ashida, H., Danchin, A. and Yokota, A. (2005) 'Was photosynthetic RuBisCO recruited by acquisitive evolution from RuBisCO-like proteins involved in sulfur metabolism?', Research in Microbiology, 156(5–6), pp. 611–618. Available at: https://doi.org/10.1016/j.resmic.2005.01.014.

Badger, M.R. et al. (1998) 'The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO2-concentrating mechanisms in algae', Canadian Journal of Botany, 76(6), pp. 1052–1071. Available at: https://doi.org/10.1139/b98-074.

Banda, D.M. et al. (2020) 'Novel bacterial clade reveals origin of form I Rubisco', Nature Plants, 6(9), pp. 1158–1166. Available at: https://doi.org/10.1038/s41477-020-00762-4.

Bar-Even, A. et al. (2011) 'The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters', Biochemistry, 50(21), pp. 4402–4410. Available at: https://doi.org/10.1021/bi2002289.

Bassham, J.A. and Krause, G.H. (1969) 'Free energy changes and metabolic regulation in steady-state photosynthetic carbon reduction', BBA - Bioenergetics, 189(2), pp. 207–221. Available at: https://doi.org/10.1016/0005-2728(69)90048-6.

Bloom, J.D. et al. (2006) ‘Structural determinants of the rate of protein evolution in yeast’, Molecular Biology and Evolution, 23(9), pp. 1751–1761. Available at: https://doi.org/10.1093/molbev/msl040.

Bouvier, J.W. et al. (2021) ‘Rubisco Adaptation Is More Limited by Phylogenetic Constraint Than by Catalytic Trade-off’, Molecular Biology and Evolution, 38(7), pp. 2880–2896. Available at: https://doi.org/10.1093/molbev/msab079.

Bouvier, J.W. and Kelly, S. (2023) ‘Response to Tcherkez and Farquhar: Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off’, Journal of Plant Physiology, 287, p. 154021. Available at: https://doi.org/10.1016/j.jplph.2023.154021.

Bowes, G., Ogren, W.L. and Hageman, R.H. (1971) Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase, Biochemical and Biophysical Research Communications. Available at: https://doi.org/10.1016/0006-291X(71)90475-X.

Buchfink, B., Reuter, K. and Drost, H.G. (2021) ‘Sensitive protein alignments at tree-of-life scale using DIAMOND’, Nature Methods, 18(4), pp. 366–368. Available at: https://doi.org/10.1038/s41592-021-01101-x.

Buchfink, B., Xie, C. and Huson, D.H. (2014) ‘Fast and sensitive protein alignment using DIAMOND’, Nature Methods, 12(1), pp. 59–60. Available at: https://doi.org/10.1038/nmeth.3176.

Busch, F.A. (2020) ‘Photorespiration in the context of Rubisco biochemistry, CO2 diffusion and metabolism’, Plant Journal, 101(4), pp. 919–939. Available at: https://doi.org/10.1111/tpj.14674.

Canfield, D.E. (2005) ‘The early history of atmospheric oxygen: Homage to Robert M. Garrels’, Annual Review of Earth and Planetary Sciences, 33, pp. 1–36. Available at: https://doi.org/10.1146/annurev.earth.33.092203.122711.

Carmo-Silva, E. et al. (2015) ‘Optimizing Rubisco and its regulation for greater resource use efficiency’, Plant, Cell and Environment, pp. 1817–1832. Available at: https://doi.org/10.1111/pce.12425.

Cavanagh, A.P., Slattery, R. and Kubien, D.S. (2022) ‘Temperature-induced changes in Arabidopsis Rubisco activity and isoform expression’, Journal of Experimental Botany [Preprint]. Available at: https://doi.org/10.1093/jxb/erac379.

Chatterjee, A. and Anindya, B. (2011) ‘Rubisco: Limitations and Re-engineering for a Better Enzyme’, International Research Journal of Plant Science, 2(2), pp. 022–024. Available at: http://interesjournals.org/IRJPS/Pdf/2011/Feb/Chatterjee and Basu.pdf (Accessed: 1 February 2022).

Chollet, R. (1977) The biochemistry of photorespiration, Trends in Biochemical Sciences. Available at: https://doi.org/10.1016/0968-0004(77)90364-4.

Christin, P.A. et al. (2008) ‘Evolutionary switch and genetic convergence on rbcL following the evolution of C4 photosynthesis’, Molecular Biology and Evolution, 25(11), pp. 2361–2368. Available at: https://doi.org/10.1093/molbev/msn178.

Cummins, P.L. (2021) ‘The Coevolution of RuBisCO, Photorespiration, and Carbon Concentrating Mechanisms in Higher Plants’, Frontiers in Plant Science, 12. Available at: https://doi.org/10.3389/fpls.2021.662425.

Cummins, P.L., Kannappan, B. and Gready, J.E. (2018) ‘Directions for optimization of photosynthetic carbon fixation: Rubisco’s efficiency may not be so constrained after all’, Frontiers in Plant Science, 9. Available at: https://doi.org/10.3389/fpls.2018.00183.

Cummins, P.L., Kannappan, B. and Gready, J.E. (2019) 'Response: Commentary: Directions for Optimization of Photosynthetic Carbon Fixation: RuBisCO's Efficiency May Not Be So Constrained After All', Frontiers in Plant Science, 10, p. 1426. Available at: https://doi.org/10.3389/fpls.2019.01426.

Dietz, K.J. and Heber, U. (1984) 'Rate-limiting factors in leaf photosynthesis. I. Carbon fluxes in the calvin cycle', BBA - Bioenergetics, 767(3), pp. 432–443. Available at: https://doi.org/10.1016/0005-2728(84)90041-0.

Drummond, D.A. et al. (2005) 'Why highly expressed proteins evolve slowly', Proceedings of the National Academy of Sciences of the United States of America, 102(40), pp. 14338–14343. Available at: https://doi.org/10.1073/pnas.0504070102.

Duraõ, P. et al. (2015) 'Opposing effects of folding and assembly chaperones on evolvability of Rubisco', Nature Chemical Biology, 11(2), pp. 148–155. Available at: https://doi.org/10.1038/nchembio.1715.

Echave, J. (2019) 'Beyond Stability Constraints: A Biophysical Model of Enzyme Evolution with Selection on Stability and Activity', Molecular Biology and Evolution, 36(3), pp. 613–620. Available at: https://doi.org/10.1093/molbev/msy244.

Eckardt, N.A. (2005) 'Photorespiration revisited', Plant Cell. American Society of Plant Biologists, pp. 2139–2141. Available at: https://doi.org/10.1105/tpc.105.035873.

Emms, D.M. and Kelly, S. (2015) 'OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy', Genome Biology, 16(1), pp. 1–14. Available at: https://doi.org/10.1186/s13059-015-0721-2.

Emms, D.M. and Kelly, S. (2019) 'OrthoFinder: Phylogenetic orthology inference for comparative genomics', Genome Biology, 20(1), pp. 1–14. Available at: https://doi.org/10.1186/s13059-019-1832-y.

Erb, T.J. and Zarzycki, J. (2018) A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2 fixing enzyme, Current Opinion in Biotechnology. Available at: https://doi.org/10.1016/j.copbio.2017.07.017.

Farquhar, G.D., von Caemmerer, S. and Berry, J.A. (1980) 'A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species', Planta, 149(1), pp. 78–90. Available at: https://doi.org/10.1007/BF00386231.

Fichtner, K. et al. (1993) 'Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with "antisense" rbcS - V. Relationship between photosynthetic rate, storage strategy, biomass allocation and vegetative plant growth at three different nitrogen s', Planta, 190(1), pp. 1–9. Available at: https://doi.org/10.1007/BF00195668.

Field, C.B. et al. (1998) 'Primary production of the biosphere: Integrating terrestrial and oceanic components', Science, 281(5374), pp. 237–240. Available at: https://doi.org/10.1126/science.281.5374.237.

Fitchen, J.H. et al. (1990) 'Residues in three conserved regions of the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase are required for quaternary structure', Proceedings of the National Academy of Sciences of the United States of America, 87(15), pp. 5768–5772. Available at: https://doi.org/10.1073/pnas.87.15.5768.

Flamholz, A. and Shih, P.M. (2020) 'Cell biology of photosynthesis over geologic time', Current Biology, 30(10), pp. R490–R494. Available at: https://doi.org/10.1016/j.cub.2020.01.076.

Flamholz, A.I. et al. (2019) ‘Revisiting Trade-offs between Rubisco Kinetic Parameters’, Biochemistry, pp. 3365–3376. Available at: https://doi.org/10.1021/acs.biochem.9b00237.

Fraser, H.B. and Hirsh, A.E. (2004) ‘Evolutionary rate depends on number of protein-protein interactions independently of gene expression level’, BMC Evolutionary Biology, 4(1), pp. 1–5. Available at: https://doi.org/10.1186/1471-2148-4-13.

Fraser, H.B., Wall, D.P. and Hirsh, A.E. (2003) ‘A simple dependence between protein evolution rate and the number of protein-protein interactions’, BMC Evolutionary Biology, 3(1), pp. 1–6. Available at: https://doi.org/10.1186/1471-2148-3-11.

Furbank, R.T. et al. (1996) ‘Antisense RNA Inhibition of RbcS Gene Expression Reduces Rubisco Level and Photosynthesis in the C4 Plant Flaveria bidentis’, Plant Physiology, 111(3), pp. 725–734. Available at: https://doi.org/10.1104/pp.111.3.725.

Gago, J. et al. (2019) ‘Photosynthesis Optimized across Land Plant Phylogeny’, Trends in Plant Science, pp. 947–958. Available at: https://doi.org/10.1016/j.tplants.2019.07.002.

Galmés, J. et al. (2005) ‘Rubisco specificity factor tends to be larger in plant species from drier habitats and in species with persistent leaves’, Plant, Cell and Environment, 28(5), pp. 571–579. Available at: https://doi.org/10.1111/j.1365-3040.2005.01300.x.

Galmés, J., Andralojc, P.J., et al. (2014) ‘Environmentally driven evolution of Rubisco and improved photosynthesis and growth within the C3 genus Limonium (Plumbaginaceae)’, New Phytologist, 203(3), pp. 989–999. Available at: https://doi.org/10.1111/nph.12858.

Galmés, J., Kapralov, M. V., et al. (2014) ‘Expanding knowledge of the Rubisco kinetics variability in plant species: Environmental and evolutionary trends’, Plant, Cell and Environment, 37(9), pp. 1989–2001. Available at: https://doi.org/10.1111/pce.12335.

Genkov, T. and Spreitzer, R.J. (2009) ‘Highly conserved small subunit residues influence Rubisco large subunit catalysis’, Journal of Biological Chemistry, 284(44), pp. 30105–30112. Available at: https://doi.org/10.1074/jbc.M109.044081.

Getzoff, T.P. et al. (1998) ‘Chimeric Arabidopsis thaliana Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Containing a Pea Small Subunit Protein Is Compromised in Carbamylation’, Plant Physiology, 116(2), pp. 695–702. Available at: https://doi.org/10.1104/pp.116.2.695.

Gielly, L. and Taberlet, P. (1994) The use of chloroplast DNA to resolve plant phylogenies: Noncoding versus rbcL sequences, Molecular Biology and Evolution. Available at: https://doi.org/10.1093/oxfordjournals.molbev.a040157.

Gout, J.F., Kahn, D. and Duret, L. (2010) ‘The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution’, PLoS Genetics, 6(5), p. 20. Available at: https://doi.org/10.1371/journal.pgen.1000944.

Grabsztunowicz, M. et al. (2015) ‘A reversible decrease in ribulose 1,5-bisphosphate carboxylase/oxygenase carboxylation activity caused by the aggregation of the enzyme’s large subunit is triggered in response to the exposure of moderate irradiance-grown plants to low irradiance’, Physiologia Plantarum, 154(4), pp. 591–608. Available at: https://doi.org/10.1111/ppl.12322.

Holland, H.D. (2002) ‘Volcanic gases, black smokers, and the great oxidation event’, Geochimica et Cosmochimica Acta, 66(21), pp. 3811–3826. Available at: https://doi.org/10.1016/S0016-7037(02)00950-X.

Horken, K.M. and Tabita, F.R. (1999) 'The "green" form I ribulose 1,5-bisphosphate carboxylase/oxygenase from the nonsulfur purple bacterium Rhodobacter capsulatus', Journal of Bacteriology, 181(13), pp. 3935–3941. Available at: https://doi.org/10.1128/jb.181.13.3935-3941.1999.

Hudson, G.S. et al. (1992) 'Reduction of ribulose-1,5-bisphosphate carboxylase/oxygenase content by antisense RNA reduces photosynthesis in transgenic tobacco plants', Plant Physiology, 98(1), pp. 294–302. Available at: https://doi.org/10.1104/pp.98.1.294.

Huson, D.H. and Scornavacca, C. (2012) 'Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks', Systematic Biology. Oxford University Press, pp. 1061–1067. Available at: https://doi.org/10.1093/sysbio/sys062.

Iñiguez, C. et al. (2020) 'Evolutionary trends in RuBisCO kinetics and their co-evolution with CO2 concentrating mechanisms', Plant Journal, 101(4), pp. 897–918. Available at: https://doi.org/10.1111/tpj.14643.

Ishikawa, C. et al. (2011) 'Functional incorporation of sorghum small subunit increases the catalytic turnover rate of rubisco in transgenic rice', Plant Physiology, 156(3), pp. 1603–1611. Available at: https://doi.org/10.1104/pp.111.177030.

Joshi, J. et al. (2015) 'Role of small subunit in mediating assembly of red-type Form I Rubisco', Journal of Biological Chemistry, 290(2), pp. 1066–1074. Available at: https://doi.org/10.1074/jbc.M114.613091.

Kanevski, I. et al. (1999) 'Plastome engineering of ribulose-1,5-bisphosphate carboxylase/oxygenase in tobacco to form a sunflower large subunit and tobacco small subunit hybrid', Plant Physiology, 119(1), pp. 133–141. Available at: https://doi.org/10.1104/pp.119.1.133.

Kapralov, M. V. et al. (2011) 'Changes in Rubisco kinetics during the evolution of C4 Photosynthesis in Flaveria (Asteraceae) are associated with positive selection on genes encoding the enzyme', Molecular Biology and Evolution, 28(4), pp. 1491–1503. Available at: https://doi.org/10.1093/molbev/msq335.

Kapralov, M. V. and Filatov, D.A. (2007) 'Widespread positive selection in the photosynthetic Rubisco enzyme', BMC Evolutionary Biology, 7. Available at: https://doi.org/10.1186/1471-2148-7-73.

Kasting, J.F. (1993) 'Earth's early atmosphere', Science, 259(5097), pp. 920–926. Available at: https://doi.org/10.1126/science.11536547.

Katoh, K. and Standley, D.M. (2013) 'MAFFT multiple sequence alignment software version 7: Improvements in performance and usability', Molecular Biology and Evolution. Oxford University Press, pp. 772–780. Available at: https://doi.org/10.1093/molbev/mst010.

Kaufman, A.J. et al. (2007) 'Late archean biospheric oxygenation and atmospheric evolution', Science, 317(5846), pp. 1900–1903. Available at: https://doi.org/10.1126/science.1138700.

Kostov, R. V., Small, C.L. and McFadden, B.A. (1997) 'Mutations in a sequence near the N-terminus of the small subunit alter the CO2/O2 specificity factor for ribulose bisphosphate carboxylase/oxygenase', Photosynthesis Research, 54(2), pp. 127–134. Available at: https://doi.org/10.1023/A:1005967106993.

Kumar, S. et al. (2017) 'TimeTree: A Resource for Timelines, Timetrees, and Divergence Times', Molecular biology and evolution, 34(7), pp. 1812–1819. Available at: https://doi.org/10.1093/molbev/msx116.

Lee, B., Berka, R.M. and Tabita, F.R. (1991) 'Mutations in the small subunit of cyanobacterial ribulose-bisphosphate carboxylase/oxygenase that modulate interactions with large subunits', Journal of Biological Chemistry, 266(12), pp. 7417–7422. Available at: https://doi.org/10.1016/s0021-9258(20)89463-0.

Lee, B., Read, B.A. and Tabita, F.R. (1991) 'Catalytic properties of recombinant octameric, hexadecameric, and heterologous cyanobacterial/ bacterial ribulose-1,5-bisphosphate carboxylase/oxygenase', Archives of Biochemistry and Biophysics, 291(2), pp. 263–269. Available at: https://doi.org/10.1016/0003-9861(91)90133-4.

Lee, B. and Tabita, F.R. (1990) 'Purification of Recombinant Ribulose-l,5-bisphosphate Carboxylase/Oxygenase Large Subunits Suitable for Reconstitution and Assembly of Active L8S8 Enzyme', Biochemistry, 29(40), pp. 9352–9357. Available at: https://doi.org/10.1021/bi00492a007.

Lin, M.T. et al. (2021) 'A procedure to introduce point mutations into the Rubisco large subunit gene in wild-type plants', Plant Journal, 106(3), pp. 876–887. Available at: https://doi.org/10.1111/tpj.15196.

Lin, M.T. et al. (2022) 'Improving the efficiency of Rubisco by resurrecting its ancestors in the family Solanaceae', Science Advances, 8(15). Available at: https://doi.org/10.1126/sciadv.abm6871.

Lobkovsky, A.E., Wolf, Y.I. and Koonin, E. V. (2010) 'Universal distribution of protein evolution rates as a consequence of protein folding physics', Proceedings of the National Academy of Sciences of the United States of America, 107(7), pp. 2983–2988. Available at: https://doi.org/10.1073/pnas.0910445107.

Van Lun, M. et al. (2014) 'CO2 and O2 distribution in rubisco suggests the small subunit functions as a CO2 reservoir', Journal of the American Chemical Society, 136(8), pp. 3165–3171. Available at: https://doi.org/10.1021/ja411579b.

Maheshwari, C. et al. (2021) 'Targeted knockdown of ribulose-1, 5-bisphosphate carboxylase-oxygenase in rice mesophyll cells', Journal of Plant Physiology, 260. Available at: https://doi.org/10.1016/j.jplph.2021.153395.

Martin-Avila, E. et al. (2020) 'Modifying plant photosynthesis and growth via simultaneous chloroplast transformation of rubisco large and small subunits', Plant Cell, 32(9), pp. 2898–2916. Available at: https://doi.org/10.1105/TPC.20.00288.

Matsumura, H. et al. (2020) 'Hybrid Rubisco with Complete Replacement of Rice Rubisco Small Subunits by Sorghum Counterparts Confers C4 Plant-like High Catalytic Activity', Molecular Plant, 13(11), pp. 1570–1581. Available at: https://doi.org/10.1016/j.molp.2020.08.012.

McFadden, B.A. and Small, C.L. (1988) 'Cloning, expression and directed mutagenesis of the genes for ribulose bisphosphate carboxylase/oxygenase', Photosynthesis Research, 18(1–2), pp. 245–260. Available at: https://doi.org/10.1007/BF00042987.

Meyer, M. and Griffiths, H. (2013) 'Origins and diversity of eukaryotic CO2-concentrating mechanisms: Lessons for the future', Journal of Experimental Botany, pp. 769–786. Available at: https://doi.org/10.1093/jxb/ers390.

Morita, K. et al. (2014) 'Unusual small subunit that is not expressed in photosynthetic cells alters the catalytic properties of Rubisco in rice', Plant Physiology, 164(1), pp. 69–79. Available at: https://doi.org/10.1104/pp.113.228015.

Nguyen, L.T. et al. (2015) 'IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies', Molecular Biology and Evolution, pp. 268–274. Available at: https://doi.org/10.1093/molbev/msu300.

Nisbet, E.G. et al. (2007) ‘The age of Rubisco: The evolution of oxygenic photosynthesis’, Geobiology, 5(4), pp. 311–335. Available at: https://doi.org/10.1111/j.1472-4669.2007.00127.x.

Nisbet, E.G. and Nisbet, R.E.R. (2008) ‘Methane, oxygen, photosynthesis, rubisco and the regulation of the air through time’, Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1504), pp. 2745–2754. Available at: https://doi.org/10.1098/rstb.2008.0057.

Ogren, W.L. (1984) ‘Photorespiration: Pathways, Regulation, and Modification’, Annual Review of Plant Physiology, 35(1), pp. 415–442. Available at: https://doi.org/10.1146/annurev.pp.35.060184.002215.

Ogren, W.L. and Bowes, G. (1971) ‘Ribulose diphosphate carboxylase regulates soybean photorespiration’, Nature New Biology, 230(13), pp. 159–160. Available at: https://doi.org/10.1038/newbio230159a0.

Orr, D.J. et al. (2017) ‘Engineering photosynthesis: Progress and perspectives’, F1000Research. Available at: https://doi.org/10.12688/f1000research.12181.1.

Orr, D.J. and Parry, M.A.J. (2020) ‘Overcoming the limitations of Rubisco: fantasy or realistic prospect?’, Journal of Plant Physiology, 254. Available at: https://doi.org/10.1016/j.jplph.2020.153285.

Pál, C., Papp, B. and Lercher, M.J. (2006) ‘An integrated view of protein evolution’, Nature Reviews Genetics, 7(5), pp. 337–348. Available at: https://doi.org/10.1038/nrg1838.

Parry, M.A.J. et al. (2003) ‘Manipulation of Rubisco: The amount, activity, function and regulation’, Journal of Experimental Botany, 54(386), pp. 1321–1333. Available at: https://doi.org/10.1093/jxb/erg141.

Parry, M.A.J. et al. (2007) ‘Prospects for increasing photosynthesis by overcoming the limitations of Rubisco’, in Journal of Agricultural Science, pp. 31–43. Available at: https://doi.org/10.1017/S0021859606006666.

Parry, M.A.J. et al. (2013) ‘Rubisco activity and regulation as targets for crop improvement’, Journal of Experimental Botany, 64(3), pp. 717–730. Available at: https://doi.org/10.1093/jxb/ers336.

Poolman, M.G., Fell, D.A. and Thomas, S. (2000) ‘Modelling photosynthesis and its control’, Journal of Experimental Botany, 51(SPEC. ISS.), pp. 319–328. Available at: https://doi.org/10.1093/jexbot/51.suppl_1.319.

Quick, W.P. et al. (1991) ‘Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with “antisense” rbcS - I. Impact on photosynthesis in ambient growth conditions’, Planta, 183(4), pp. 542–554. Available at: https://doi.org/10.1007/BF00194276.

Read, B.A. and Tabita, F.R. (1992a) ‘A Hybrid Ribulosebisphosphate Carboxylase/Oxygenase Enzyme Exhibiting a Substantial Increase in Substrate Specificity Factor’, Biochemistry, 31(24), pp. 5553–5560. Available at: https://doi.org/10.1021/bi00139a018.

Read, B.A. and Tabita, F.R. (1992b) ‘Amino Acid Substitutions in the Small Subunit of Ribulose-1, 5-bisphosphate Carboxylase/Oxygenase That Influence Catalytic Activity of the Holoenzyme’, Biochemistry, 31(2), pp. 519–525. Available at: https://doi.org/10.1021/bi00117a031.

Rickaby, R.E.M. and Eason Hubbard, M.R. (2019) ‘Upper ocean oxygenation, evolution of RuBisCO and the Phanerozoic succession of phytoplankton’, Free Radical Biology and Medicine, 140, pp. 295–304. Available at: https://doi.org/10.1016/j.freeradbiomed.2019.05.006.

Robbins, E.H.J. and Kelly, S. (2023) 'The Evolutionary Constraints on Angiosperm Chloroplast Adaptation', Genome Biology and Evolution, 15(6). Available at: https://doi.org/10.1093/gbe/evad101.

Saeed, R. and Deane, C.M. (2006) 'Protein protein interactions, evolutionary rate, abundance and age', BMC Bioinformatics, 7. Available at: https://doi.org/10.1186/1471-2105-7-128.

Sakoda, K. et al. (2021) 'Effects of introduction of sorghum RbcS with rice RbcS knockdown by RNAi on photosynthetic activity and dry weight in rice', Plant Production Science, 24(3), pp. 346–353. Available at: https://doi.org/10.1080/1343943X.2020.1847669.

Sarles, L.S. and Tabita, F.R. (1983) 'Derepression of the synthesis of D-ribulose 1,5-bisphosphate carboxylase/oxygenase from Rhodospirillum rubrum', Journal of Bacteriology, 153(1), pp. 458–464. Available at: https://doi.org/10.1128/jb.153.1.458-464.1983.

Savir, Y. et al. (2010) 'Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape', Proceedings of the National Academy of Sciences of the United States of America, 107(8), pp. 3475–3480. Available at: https://doi.org/10.1073/pnas.0911663107.

Schlüter, U. et al. (2023) 'Brassicaceae display variation in efficiency of photorespiratory carbon recapturing mechanisms', Journal of Experimental Botany [Preprint]. Available at: https://doi.org/10.1093/jxb/erad250.

Schneider, G. et al. (1990) 'Comparison of the crystal structures of L2 and L8S8 Rubisco suggests a functional role for the small subunit', EMBO Journal, 9(7), pp. 2045–2050. Available at: https://doi.org/10.1002/j.1460-2075.1990.tb07371.x.

Schneider, G., Lindqvist, Y. and Brändén, C.I. (1992) 'RUBISCO: Structure and mechanism', Annual Review of Biophysics and Biomolecular Structure, pp. 119–143. Available at: https://doi.org/10.1146/annurev.bb.21.060192.001003.

Schulz, L. et al. (2022) 'Evolution of increased complexity and specificity at the dawn of form I Rubiscos', Science, 378(6616). Available at: https://doi.org/10.1126/science.abq1416.

Seward, E.A. and Kelly, S. (2018) 'Selection-driven cost-efficiency optimization of transcripts modulates gene evolutionary rate in bacteria', Genome Biology, 19(1). Available at: https://doi.org/10.1186/s13059-018-1480-7.

Sharkey, T.D. (2020) 'Emerging research in plant photosynthesis', Emerging Topics in Life Sciences, 4(2), pp. 137–150. Available at: https://doi.org/10.1042/ETLS20200035.

Sharwood, R.E. (2017) 'Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops', New Phytologist, pp. 494–510. Available at: https://doi.org/10.1111/nph.14351.

Sharwood, R.E., Ghannoum, O. and Whitney, S.M. (2016) 'Prospects for improving CO2 fixation in C3-crops through understanding C4-Rubisco biogenesis and catalytic diversity', Current Opinion in Plant Biology, pp. 135–142. Available at: https://doi.org/10.1016/j.pbi.2016.04.002.

Shih, P.M. et al. (2016) 'Biochemical characterization of predicted Precambrian RuBisCO', Nature Communications, 7. Available at: https://doi.org/10.1038/ncomms10382.

Ślesak, I., Ślesak, H. and Kruk, J. (2017) 'RubisCO Early Oxygenase Activity: A Kinetic and Evolutionary Perspective', BioEssays, 39(11). Available at: https://doi.org/10.1002/bies.201700071.

Soskine, M. and Tawfik, D.S. (2010) 'Mutational effects and the evolution of new protein functions', Nature Reviews Genetics, 11(8), pp. 572–582. Available at: https://doi.org/10.1038/nrg2808.

Spreitzer, R.J. (2003) 'Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase', Archives of Biochemistry and Biophysics, 414(2), pp. 141–149. Available at: https://doi.org/10.1016/S0003-9861(03)00171-1.

Stitt, M. et al. (1991) 'Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with "antisense" rbcS - II. Flux-control coefficients for photosynthesis in varying light, CO2, and air humidity', Planta, 183(4), pp. 555–566. Available at: https://doi.org/10.1007/BF00194277.

Studer, R.A. et al. (2014) 'Stability-activity tradeoffs constrain the adaptive evolution of RubisCO', Proceedings of the National Academy of Sciences of the United States of America, 111(6), pp. 2223–2228. Available at: https://doi.org/10.1073/pnas.1310811111.

Su, Z. and Townsend, J.P. (2015) 'Utility of characters evolving at diverse rates of evolution to resolve quartet trees with unequal branch lengths: Analytical predictions of long-branch effects', BMC Evolutionary Biology, p. 86. Available at: https://doi.org/10.1186/s12862-015-0364-7.

Suyama, M., Torrents, D. and Bork, P. (2006) 'PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments', Nucleic Acids Research, 34(WEB. SERV. ISS.), p. W609. Available at: https://doi.org/10.1093/nar/gkl315.

Tabita, F.R. (1999) 'Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective', Photosynthesis Research, 60(1), pp. 1–28. Available at: https://doi.org/10.1023/A:1006211417981.

Tabita, F.R., Satagopan, S., et al. (2008) 'Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships', Journal of Experimental Botany, 59(7), pp. 1515–1524. Available at: https://doi.org/10.1093/jxb/erm361.

Tabita, F.R., Hanson, T.E., et al. (2008) 'Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms', Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1504), pp. 2629–2640. Available at: https://doi.org/10.1098/rstb.2008.0023.

Tcherkez, G., Farquhar, G.D., 2021. Rubisco catalytic adaptation is mostly driven by photosynthetic conditions – Not by phylogenetic constraints. J Plant Physiol 267. https://doi.org/10.1016/j.jplph.2021.153554

Tcherkez, G.G.B., Farquhar, G.D. and Andrews, T.J. (2006) 'Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized', Proceedings of the National Academy of Sciences of the United States of America, 103(19), pp. 7246–7251. Available at: https://doi.org/10.1073/pnas.0600605103.

Tóth-Petróczy, Á. and Tawfik, D.S. (2011) 'Slow protein evolutionary rates are dictated by surface - core association', Proceedings of the National Academy of Sciences of the United States of America, 108(27), pp. 11151–11156. Available at: https://doi.org/10.1073/pnas.1015994108.

Tourasse, N.J. and Li, W.H. (2000) 'Selective constraints, amino acid composition, and the rate of protein evolution', Molecular Biology and Evolution, 17(4), pp. 656–664. Available at: https://doi.org/10.1093/oxfordjournals.molbev.a026344.

van der Vies, S.M., Bradley, D. and Gatenby, A.A. (1986) ' Assembly of cyanobacterial and higher plant ribulose bisphosphate carboxylase subunits into functional homologous and heterologous enzyme molecules in Escherichia coli ', The EMBO Journal, 5(10), pp. 2439–2444. Available at: https://doi.org/10.1002/j.1460-2075.1986.tb04519.x.

Voordouw, G. et al. (1987) ‘Site-directed mutagenesis of the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from Anacystis nidulans’, European Journal of Biochemistry, 163(3), pp. 591–598. Available at: https://doi.org/10.1111/j.1432-1033.1987.tb10908.x.

Wall, D.P. et al. (2005) ‘Functional genomic analysis of the rates of protein evolution’, Proceedings of the National Academy of Sciences of the United States of America, 102(15), pp. 5483–5488. Available at: https://doi.org/10.1073/pnas.0501761102.

Wang, Y.L. et al. (2001) ‘Properties of hybrid enzymes between Synechococcus large subunits and higher plant small subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase in Escherichia coli’, Archives of Biochemistry and Biophysics, 396(1), pp. 35–42. Available at: https://doi.org/10.1006/abbi.2001.2555.

Whitney, S.M. et al. (2011) ‘Isoleucine 309 acts as a C 4 catalytic switch that increases ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) carboxylation rate in flaveria’, Proceedings of the National Academy of Sciences of the United States of America, 108(35), pp. 14688–14693. Available at: https://doi.org/10.1073/PNAS.1109503108/-/DCSUPPLEMENTAL.

Whitney, S.M. et al. (2015) ‘Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone’, Proceedings of the National Academy of Sciences of the United States of America, 112(11), pp. 3564–3569. Available at: https://doi.org/10.1073/pnas.1420536112.

Whitney, S.M., Houtz, R.L. and Alonso, H. (2011) ‘Advancing our understanding and capacity to engineer nature’s CO2-sequestering enzyme, Rubisco’, Plant Physiology, 155(1), pp. 27–35. Available at: https://doi.org/10.1104/pp.110.164814.

Zhou, Y. et al. (2023) ‘Grafting Rhodobacter sphaeroides with red algae Rubisco to accelerate catalysis and plant growth’, Nature Plants, 9(6), pp. 978–986. Available at: https://doi.org/10.1038/s41477-023-01436-7.

Zhu, X.G., De Sturler, E. and Long, S.P. (2007) ‘Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: A numerical simulation using an evolutionary algorithm’, Plant Physiology, 145(2), pp. 513–526. Available at: https://doi.org/10.1104/pp.107.103713.

Chapter 4. Metabolic engineering of stomatal precursor cells enhances photosynthetic water-use efficiency and vegetative growth under water-deficit conditions in Arabidopsis thaliana

Chapter 4, Contributors:

Bouvier, J.W. and Kelly, S.

Chapter 4, Author contributions:

J.W.B. and S.K. conceived the study, designed the experiments, interpreted and analyzed the data, and wrote the manuscript. J.W.B. conducted the experiments.

This thesis chapter has been published as pre-print at the following location:

Bouvier, J.W. and Kelly, S. 2024. bioRxiv, doi 10.1101/2024.08.29.610053

All supplemental data files are provided in the online version of this manuscript.

4.1 Abstract

Stomata are epidermal pores that control the exchange of gaseous CO_2 and H_2O between plants and their environment. Modulating stomatal density can alter this exchange, and thus presents a target for engineering improved crop productivity and climate resilience. Here, we show that stomatal density in *Arabidopsis thaliana* can be decreased by the expression of a water-forming NAD(P)H oxidase targeted to stomatal precursor cells. We demonstrate that this reduction in stomatal density occurs irrespective of whether the expressed enzyme is targeted to the cytosol, chloroplast stroma,

or chloroplast intermembrane space of these cells. We reveal that this decrease in stomatal density occurs in the absence of any measurable impact on stomatal dynamics, or the efficiency or thermal sensitivity of photosynthesis. Consequently, overexpression plants exhibit a higher intrinsic water-use efficiency due to an increase in CO_2 fixed per unit water transpired. Finally, we demonstrate that this enhanced water-use efficiency translates to an improvement in vegetative growth and biomass accumulation under water-deficit conditions. Together, these results thus provide a novel approach for enhancing plant productivity through metabolic engineering of stomatal density.

4.2 Introduction

Stomata mediate the exchange of gaseous CO_2 and water vapour between plants and their external environment. Specifically, CO_2 provides the primary substrate of photosynthesis and its uptake is thereby required to fuel plant growth. In contrast, loss of water vapour is integral to the transpiration stream which helps regulate internal water status, temperature, and nutrient uptake. As both CO_2 and H_2O gases share a single diffusion path but in opposite directions, controlling flux along this path is essential for balancing the physiological demands of the plant. Consequently, optimizing the distribution and behaviour of stomata is a central way in which plants adapt to different terrestrial habitats (Edwards, Kerp and Hass, 1998; Raven, 2002; Clark *et al.*, 2022).

Given their role at the interface between plants and their environment, plants are able to modulate the diffusive conductance of stomata in response to diverse stimuli including light (quality and quantity), humidity, temperature, soil water availability, and atmospheric CO_2 (Driesen *et al.,* 2020). By integrating these signals alongside natural circadian rhythms, stomata open to promote CO_2 fixation during the day and close at night or under adverse conditions to limit excessive water loss. In addition to this complex and dynamic behaviour, stomata also exhibit remarkable developmental plasticity across species and in response to diverse environmental changes. Most notably, stomatal density inversely correlates with soil water availability (Quarrie and Jones, 1977), temperature (Lau *et al.*, 2018), and ambient CO_2 concentration (Woodward, 1987; Woodward and Bazzaz, 1988; McElwain and Chaloner, 1995; Beerling, McElwain and Osborne, 1998; Beerling, Osborne and Chaloner, 2001; Woodward, Lake and Quick, 2002; Franks and Beerling, 2009), and positively

correlates with relative air humidity (Bakker, 1991; Fanourakis, Heuvelink and Carvalho, 2013) over both developmental or geological timescales. These environmentally-induced changes in density are generally linked with secondary adjustments in stomatal size (such that higher densities of stomata are generally smaller and *vice versa* (Doheny-Adams *et al.*, 2012; de Boer *et al.*, 2016; Dittberner *et al.*, 2018; Buckley, Caine and Gray, 2020)), though density always takes precedence over size as the predominant anatomical constraint on leaf gaseous conductance (McElwain and Chaloner, 1995; Mcelwain, Yiotis and Lawson, 2016). Thus, both dynamic and developmental changes in stomata play important roles in the growth and environmental adaptation of plants.

The considerable plasticity of stomatal form and function, coupled with the importance of stomata in regulating plant growth and environmental interactions, have together inspired multiple attempts to alter their properties for crop improvement. Pioneering efforts of Farquhar and Richards in the 1980's used stable carbon isotope ratios to screen for enhanced water use efficiency in wheat (Farquhar and Richards, 1984; Condon *et al.*, 2002). Subsequently, targeted engineering efforts have included a range of forward genetic approaches which have manipulated the distribution of stomata (Nadeau and Sack, 2002; Shpak *et al.*, 2005; Wang *et al.*, 2007; Dow, Berry and Bergmann, 2014; Papanatsiou, Amtmann and Blatt, 2016; Yu *et al.*, 2020), as well as the mechanical (Carroll *et al.*, 2022), transport (Kusumi *et al.*, 2012; Wang *et al.*, 2014; Medeiros *et al.*, 2016, 2017; Antunes *et al.*, 2017; Papanatsiou *et al.*, 2019; Wang *et al.*, 2019; Flütsch, Nigro, *et al.*, 2020; Reyer *et al.*, 2020; Zhou *et al.*, 2021; Huang *et al.*, 2021; Toh *et al.*, 2021; Horaruang *et al.*, 2022), and metabolic (Laporte, Shen and Tarczynski, 2002; Antunes *et al.*, 2012; Penfield *et al.*, 2012; Antunes *et al.*, 2017; Kelly *et al.*, 2012, 2019; Lugassi *et al.*, 2015, 2019; Prasch *et al.*, 2015; Daloso *et al.*, 2016; Horrer *et al.*, 2016; Müller *et al.*, 2018; Boxall *et al.*, 2020; Flütsch, Wang, *et al.*, 2020; Acevedo-Siaca *et al.*, 2022) properties of their guard cells. Of all these approaches, altering stomatal density has received the most attention, and a vast array of mutants with both increased (Yang and Sack, 1995; Schlüter *et al.*, 2003; Bergmann, Lukowitz and Somerville, 2004; Masle, Gilmore and Farquhar, 2005; Bussis *et al.*, 2006; Pillitteri *et al.*, 2007; Gudesblat *et al.*, 2012; Yu Tanaka *et al.*, 2013; Franks *et al.*, 2015; Hepworth *et al.*, 2015; Schuler *et al.*, 2018; Jalakas *et al.*, 2018; Sakoda *et al.*, 2020) and decreased (Bergmann, Lukowitz and Somerville, 2004; Bussis *et al.*, 2006; Pillitteri

et al., 2007; Hara *et al.*, 2007, 2009; Yoo *et al.*, 2010; Doheny-Adams *et al.*, 2012; Gudesblat *et al.*, 2012; Franks *et al.*, 2015; Hepworth *et al.*, 2015; Wang *et al.*, 2016; Hughes *et al.*, 2017; Yin *et al.*, 2017; Jalakas *et al.*, 2018; Mohammed *et al.*, 2019; Caine *et al.*, 2019; Dunn *et al.*, 2019; Zhao *et al.*, 2024) stomatal densities have been engineered across diverse plant and crop species.

A common theme from approaches which have altered stomatal densities has been the genetic manipulation of the expression levels of stomatal development genes. Such targets have included the basic helix-loop-helix transcription factors SPEECHLESS (SPCH) (Gudesblat *et al.*, 2012), MUTE (Pillitteri *et al.*, 2007), and FAMA (Bergmann, Lukowitz and Somerville, 2004) which regulate the initiation, proliferation, and differentiation stages of the stomatal lineage, respectively. Engineering success has also been achieved by altering the expression levels of other important components of stomatal development including the EPIDERMAL PATTERNING FACTOR family of signalling peptides (Hara *et al.*, 2007, 2009; Doheny-Adams *et al.*, 2012; Yu Tanaka *et al.*, 2013; Franks *et al.*, 2015; Hepworth *et al.*, 2015; Wang *et al.*, 2016; Yin *et al.*, 2017; Hughes *et al.*, 2017; Caine *et al.*, 2019; Dunn *et al.*, 2019; Mohammed *et al.*, 2019; Sakoda *et al.*, 2020) alongside their receptor components (TOO MANY MOUTHS (Yang and Sack, 1995) and the ERECTA protein kinases (Masle, Gilmore and Farquhar, 2005)), stomatal density and distribution 1 (SDD1) and its respective interactor components (Schlüter *et al.*, 2003; Bussis *et al.*, 2006; Yoo *et al.*, 2010), and a host of other genes involved in hormone signalling and non-canonical stomatal developmental pathways (Jalakas *et al.*, 2018; Schuler *et al.*, 2018; Zhao *et al.*, 2024). However, despite these considerable successes, the molecular and biochemical mechanisms that link these gene networks to changes in environmental and physiological cues remain poorly understood.

Nicotinamide adenine dinucleotide (NAD^+) presents a potential metabolic nexus that can link physiological changes in the leaf to the genetic regulation of stomatal patterning (de Souza Chaves *et al.*, 2019; Feitosa-Araujo, da Fonseca-Pereira, *et al.*, 2020; Falquetto-Gomes *et al.*, 2024). This potential role arises from recent studies which have demonstrated that changes in cellular NAD^+ status either through exogenous NAD^+ treatment (Feitosa-Araujo, da Fonseca-Pereira, *et al.*, 2020), knockdown of mitochondrial (NDT1 and NDT2) and peroxisomal (PXN1) NAD^+ transporters (de Souza Chaves *et al.*, 2019; Feitosa-Araujo, da Fonseca-Pereira, *et al.*, 2020), or knockdown of

poly(ADP-ribose)polymerase 1 gene involved in NAD^+ recycling (Feitosa-Araujo, da Fonseca-Pereira, *et al.*, 2020) all produce plants with decreased stomatal densities in *Arabidopsis thaliana*. This NAD^+-mediated alteration in stomatal patterning is thought to occur via perturbation of abscisic acid (ABA) biosynthesis and signalling, which supresses the activity of SPCH (Yang *et al.*, 2022) and in turn acts to inhibit cell entry into the stomatal lineage (Quarrie and Jones, 1977; Lake and Woodward, 2008; Yoko Tanaka *et al.*, 2013; Chater *et al.*, 2014, 2015). This connection between NAD^+ and ABA was postulated because changes in NAD^+ homeostasis were associated with both transcriptional changes in ABA metabolic genes and endogenous levels of this phytohormone, and because alterations in stomatal density in mutants could be rescued by inhibition of ABA biosynthesis (Feitosa-Araujo, da Fonseca-Pereira, *et al.*, 2020). Intriguingly, this also aligns with the known interaction of both NAD^+ and ABA in regulating stomatal aperture (Hashida *et al.*, 2010; Sun *et al.*, 2017). Thus, as NAD^+ is one of the cornerstones of plant metabolism, the recent discovery that it also has a role in regulating stomatal cell differentiation provides a simple and unified route through which changes in physiology or environment directly feedback on stomatal pattering.

Water-forming NAD(P)H oxidases (NOX) present a powerful genetically-encoded tool for studying the role of redox status on organismal developmental processes. Specifically, NOX are a class of enzymes which catalyse the four-electron reduction of molecular oxygen in the presence of protons and reducing equivalents NADH and nicotinamide adenine dinucleotide phosphate (NADPH) to produce water and the corresponding $NAD^+/NADP^+$ electron acceptor in the following stoichiometry [$2NAD(P)H + 4H^+ + O_2 \rightarrow 2NADP^+ + 2H_2O$]. In nature, these enzymes are thought to function in both oxidative stress management as well as in balancing the relative levels of oxidised and reduced forms of co-factor pools (Yan *et al.*, 2016). Due to this physiological role and the fact that NOX enzymes are highly catalytically efficient, they have been exploited to alter ratios of oxidised and reduced co-factor pools in both fundamental and applied science contexts (Kroutil *et al.*, 2004; Eiteman *et al.*, 2006; Heux, Cachon and Dequin, 2006; Vemuri *et al.*, 2007; Turner, 2011; de Gonzalo, A. Orden and R. Bisogno, 2012; Titov *et al.*, 2016; Cracan *et al.*, 2017).

In this study, we aimed to engineer stomatal density through NOX-mediated manipulation of $NAD(P)^+/NAD(P)H$ status in stomatal precursor cells. We show that active monomeric NOX can be

targeted to different compartments in plant cells. We further show that expression of NOX across these diverse subcellular compartments of stomatal precursor cells leads to a reduction in stomatal density without any adverse effects on stomatal function or leaf-level photosynthesis. Although maximal rates of CO_2 assimilation are not affected in transgenic plants, the intrinsic photosynthetic water-use efficiency is improved such that more CO_2 is fixed per unit water transpired by the leaf. Accordingly, this results in enhanced growth and biomass accumulation under water-deficit conditions, whilst maintaining wild-type levels of growth under well-watered conditions. In summary, this study thus provides a novel and tractable metabolic engineering approach to manipulate stomatal patterning as a potential target to enhance crop productivity and climate resilience.

4.3 Results

4.3.1 NOX can be expressed in plant cells and targeted to multiple subcellular locations

Diverse water-forming NAD(P)H oxidases (NOX) enzymes have been characterised across bacteria (Schmidt *et al.*, 1986; Stanton and Jensen, 1993; Higuchi *et al.*, 1993; Matsumoto *et al.*, 1996; Sakamoto, Uchimura and Komagata, 1996; Lopez de Felipe and Hugenholtz, 2001; Riebel *et al.*, 2002; Geueke, Riebel and Hummel, 2003; Hummel and Riebel, 2003; Kawasaki *et al.*, 2004; Lountos *et al.*, 2006; Park *et al.*, 2011; Wang, Chong and Jiang, 2012; Gao *et al.*, 2012; Zhang *et al.*, 2012; Gao *et al.*, 2019; Nowak *et al.*, 2015; Titov *et al.*, 2016; Cracan *et al.*, 2017; Dishisha *et al.*, 2019; Li *et al.*, 2019), archaea (Ward *et al.*, 2001; Jia *et al.*, 2008, 2010; Rashid *et al.*, 2011; Wu *et al.*, 2012; Nisar *et al.*, 2013; Yan *et al.*, 2016), and eukaryotes (Brown, Upcroft and Upcroft, 1996; Li and Wang, 2006; Castillo-Villanueva *et al.*, 2016). Here, we used an engineered variant of this enzyme from the bacterium *Streptococcus mutans* (*Sm*NOX) for the basis of all experiments. This variant was chosen based on several criteria. First, *Sm*NOX exhibits NAD(P)H bifunctionality (Petschacher *et al.*, 2014) and is thus capable of oxidising co-factor pools in a range of eukaryotic cellular contexts. Second, *Sm*NOX is functional across a broad pH and temperature range with optimal activities matching physiological conditions *in planta* (Higuchi *et al.*, 1993; Li *et al.*, 2018). Third, *Sm*NOX is monomeric (Higuchi *et al.*, 1993) allowing a higher possibility for it to tolerate protein-protein fusions and N-

terminal targeting peptides whilst maintaining functional activity. Finally, *Sm*NOX exhibits negligible secondary H_2O_2 production that may contribute to undesirable off-target effects (Petschacher ***et al.***, 2014; Li ***et al.***, 2018).

To assess whether NOX can be produced in plants, we first tested *Sm*NOX expression using a transient protoplast expression system. For this purpose, mesophyll protoplasts were isolated from mature rosette leaves of *A. thaliana* and transformed with a genetic construct containing the full-length *Sm*NOX coding sequence translationally fused to a C-terminal GFP (Appendix D [and ***Supplemental File 1 online***], Figure S1A). Analysis of cell fluorescence using confocal microscopy confirmed that the *Sm*NOX-GFP fusion was translated in plant cells, and that the enzyme localised to the cytosol when driven in the absence of any additional targeting sequences (Figure 4.1A) in agreement with a predicted bioinformatic localisation (Appendix D [and ***Supplemental File 1 online***], table S1).

Whilst NAD^+ is produced in the cytosol, the regulatory mechanism that governs cytosolic NAD^+ levels is dependent on chloroplast $NADP^+$ production. Accordingly, we sought to test whether *Sm*NOX could also be localised closer to the site of $NADP^+$ supply in both the chloroplast stroma (the site of NADPH production) and the chloroplast intermembrane space (the site of NADPH/NADH transfer via the malate valve). *Sm*NOX was successfully localised to the chloroplast stroma (stromal-*Sm*NOX) when fused to the transit peptide from the photosystem II subunit QA (PSBQA) protein (Figure 4.1A). This stromal localisation was verified by comparison to a stromal control consisting of the full-length rubisco small subunit gene fused to GFP (Figure 4.1A). Moreover, *Sm*NOX was also successfully targeted to the chloroplast intermembrane space (CIMS-*Sm*NOX) when fused to the full-length outer envelope protein 9 gene (Figure 4.1A). In summary, *Sm*NOX can be expressed in plants and can be localised to diverse subcellular compartments including the cytosol (the site of NAD^+ production), chloroplast stroma (the site of $NADP^+$ production), and chloroplast intermembrane space (the site of NAD^+/$NADP^+$ transfer).

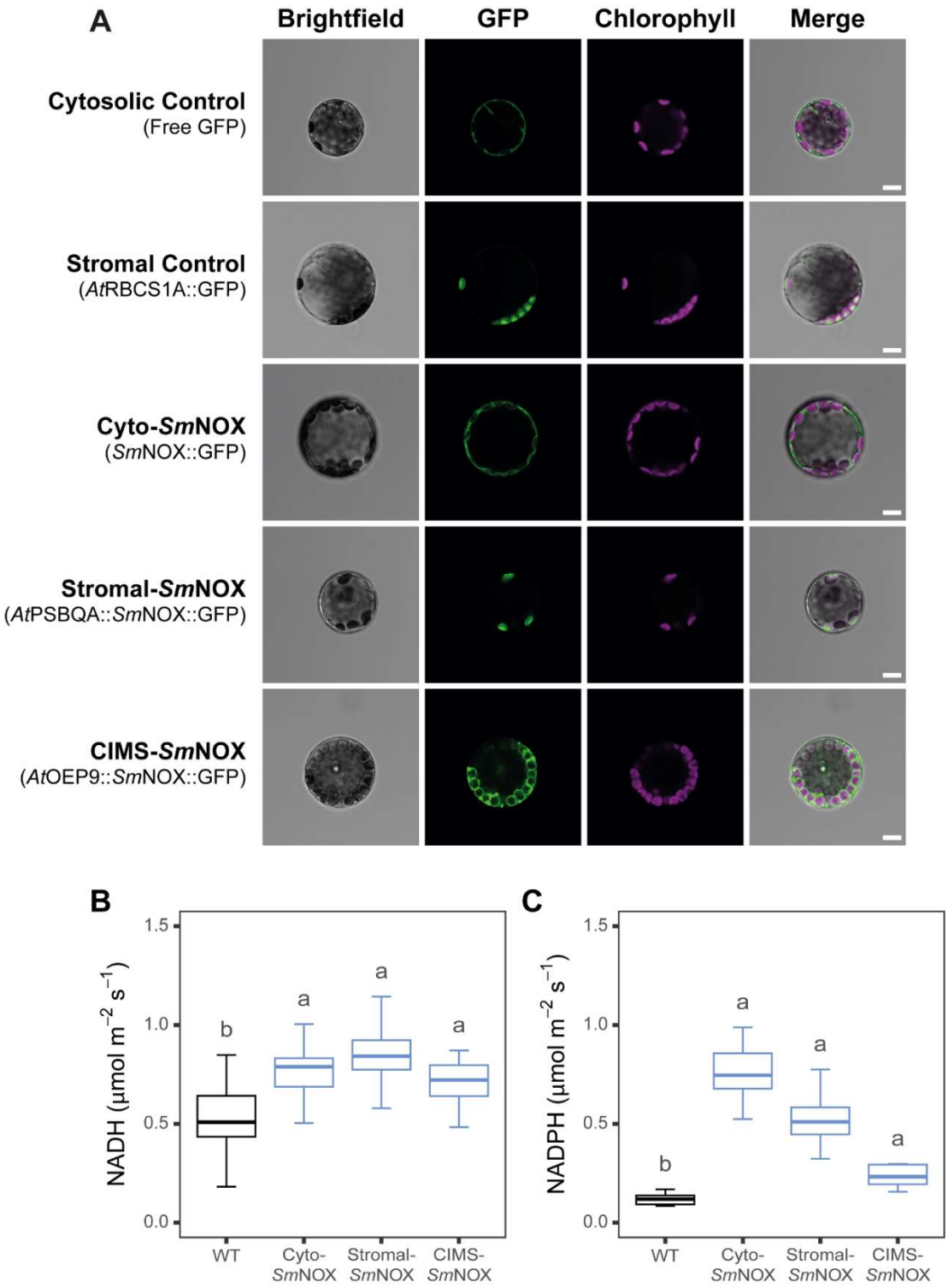


Figure 4.1. Validation of *Sm*NOX localisation and functional activity *in planta*. A) The subcellular localisation of *Sm*NOX in *Arabidopsis thaliana* protoplasts. Free GFP: protoplasts expressing eGFP (cytosolic control). *At*RBCS1A::GFP: protoplasts expressing eGFP fused to the C-terminus of the rubisco small subunit (chloroplast stroma control). Cyto-*Sm*NOX: protoplasts expressing *Sm*NOX fused at the C-terminus to eGFP (cytosolic localisation). Stromal-*SmNOX*: protoplasts expressing *SmNOX* fused at the C-terminus to eGFP and fused at the N-terminus to the chloroplast transit peptide of the photosystem II subunit QA protein (chloroplast stromal localisation). CIMS-*Sm*NOX: protoplasts expressing *SmNOX* fused at the C-terminus to eGFP and fused at the

N-terminus to the full length outer envelope protein 9 (CIMS; chloroplast intermembrane space localisation). Three separate signals are shown. Green: eGFP fluorescence. Violet: Thylakoid-derived chlorophyll autofluorescence. Grey: bright-field. The merged channel represents all three signals overlaid in a single image. Scale bar = 10 µM. B) Boxplot depicting the rate of NADH oxidation from crude leaf lysate of wild-type (black) and transgenic plants (blue). Abbreviations follow that described in (A). Data represent an average across three independent single copy lines (*n* = 4 plants per line). C) As in (B), but for the total NADPH oxidation from crude leaf lysate (*n* = 5 plants per line). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). The raw data can be found in ***Supplemental File 4 online***. **[*End of Figure Legend*].**

4.3.2 Identifying a promoter capable of driving stable and high expression of SmNOX in stomatal precursor cells

To investigate whether *Sm*NOX could affect plant stomatal patterning, it was necessary to next find a promoter which drives high gene expression in undifferentiated stomatal progenitor cells in the leaf epidermis (i.e., during the developmental window prior to onset of stomatal lineage differentiation) (Appendix D [and ***Supplemental File 1 online***], Figure S2A). For this purpose, we analysed a cell-specific transcriptome dataset derived from all major cell types spanning the stomatal developmental trajectory (Adrian *et al.*, 2015). This revealed that the expression of genes encoding known master-regulators of stomatal development including the transcription factors SPEECHLESS (SPCH), MUTE, and FAMA were expressed at low levels which were considered to be insufficient to achieve *Sm*NOX-mediated alteration of cellular NAD^+ status (Appendix D [and ***Supplemental File 1 online***], Figure S2B – S2J). Consequently, we sought to identify an alternative more highly expressed promoter.

Interrogation of the above transcriptome data revealed that the gene encoding the chlorophyll a/b binding protein 3 (CAB3) exhibited high mRNA abundance in the target epidermal tissue of interest (Appendix D [and ***Supplemental File 1 online***], Figure S2K). Specifically, CAB3 was ranked as the 36th (out of ~28,000) most highly expressed nuclear-encoded gene in this cell type (Appendix D [and ***Supplemental File 1 online***], Figure S2L and S2M), and exhibited transcripts levels that were

~1,800-fold more abundant than SPCH (0.5 TPM), ~9,000-fold more abundant than MUTE (0.1 TPM), and ~425-fold more abundant than FAMA (2.1 TPM), respectively (Appendix D [and ***Supplemental File 1 online***], Figure S2B – S2M). Thus, as CAB3 is one of the most abundant transcripts in the cell type of interest, and as CAB3 is a well-characterised promoter that is extensively used for driving transgene expression in plant synthetic biology (Mitra, Choi and An, 1989), this promoter was selected for use herein. Genetic constructs containing targeted and untargeted variants of the *Sm*NOX coding sequence were thus cloned under the control of the CAB3 promoter and were transformed into *A. thaliana* ecotype Col-0 (Appendix D [and ***Supplemental File 1 online***], Figure S1B and Figure S3A and S3B). Homozygous plants from three single-copy insertion events of cyto-SmNOX, stromal-SmNOX and CIMS-SmNOX were used for the basis of all subsequent experiments.

4.3.3 Validation of SmNOX expression and function

To assess the functional activity of NOX in plants, total rates of NADH and NADPH oxidation were determined using *in vitro* assays containing crude lysate from mature rosette leaves. As plants possess multiple native enzymes capable of oxidising NADH and NADPH, background oxidation of NADH and NADPH was observed in all cases as expected (Figure 4.1B and 4.1C). However, the rates of both NADH and NADPH oxidation were significantly increased across all transgenic lines compared to wild-type controls (Figure 4.1B and 4.1C). Thus, there was successful expression of active *Sm*NOX in each of the targeted subcellular localisations.

It should be noted that the CAB3 promoter also drives high levels of transgene expression in other photosynthetic cells of the mature leaf (Appendix D [and ***Supplemental File 1 online***], Figure S2K – S2M and Figure S4A – S4C). Thus, although our target tissue for expression of *Sm*NOX is the leaf epidermis, only a portion of the above enzyme activity in the crude leaf lysate corresponds to this cell layer. Moreover, the rates of *Sm*NOX-mediated NADH/NADPH oxidation measured above (~ 1.0 µmol m^{-2} s^{-1}) are substantial in the context of an epidermal cell, they are not expected to impact NADH/NADPH pools in photosynthetic mesophyll cells. This is because of both the large pool size of NADH (24 – 55 µM) and NADPH (140 – 180 µM) (Igamberdiev and Gardeström, 2003) in the mesophyll tissue, as well as the extraordinary high rates of photosynthetic NADPH production (~70

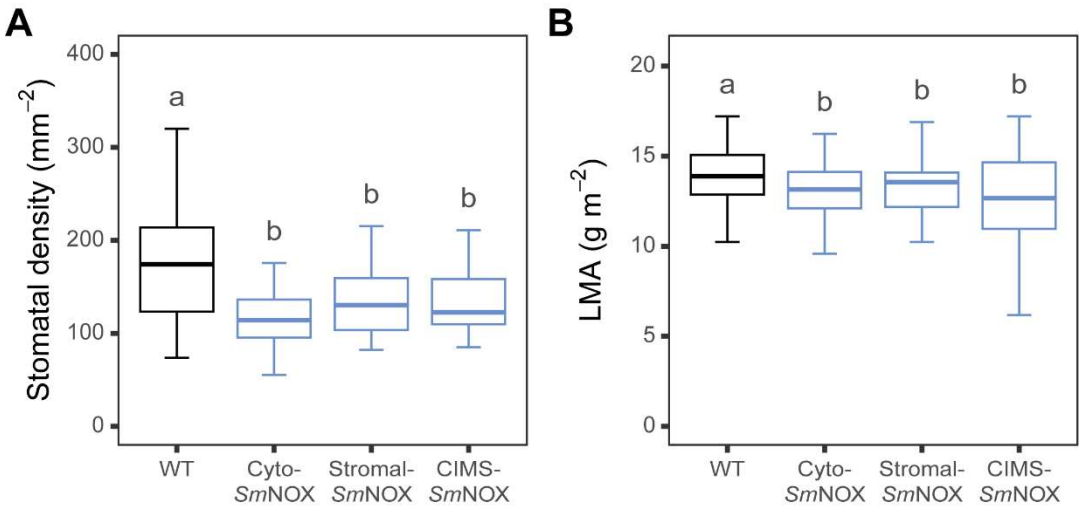


Figure 4.2. The effect of *Sm*NOX expression on stomatal density. A) Boxplot depicting the stomatal density (count mm^{-2}) measured on the abaxial surface of mature rosette leaves from wild-type (black) and transgenic plants (blue). WT: wild-type. Cyto-*Sm*NOX: cytosolic *Sm*NOX. Stromal-*Sm*NOX: chloroplast stroma *Sm*NOX. CIMS-*Sm*NOX: chloroplast intermembrane space *Sm*NOX. Data represent an average across three independent single copy lines (*n* = 9 – 10 plants per line). B) As in (A), but for the measured leaf mass per area (LMA, g dry mass per m^{-2} leaf area) of mature rosette leaves (*n* = 10 plants per line). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). The raw data can be found in ***Supplemental File 4 online***. **[*End of Figure Legend*].**

µmol g^{-1} s^{-1}) (Wieloch and Sharkey, 2022) which occurs during the light. Thus, we hypothesized that *Sm*NOX would be capable of manipulating endogenous levels of NAD$^+$ in leaf cells with low photosynthetic activity such as epidermal cells, but would have little or no effect on the NAD$^+$ status or photosynthesis of mesophyll cells.

4.3.4 SmNOX activity causes reductions in stomatal density

Given that *Sm*NOX was active in the transgenic plants, we next investigated whether this resulted in an alteration in stomatal patterning. For this purpose, the epidermal composition of the abaxial surface of mature rosette leaves from 8-week-old plants was analysed by bright-field microscopy. This analysis revealed that stomatal densities in all *Sm*NOX-expressing lines were reduced by 20 – 30% compared to wild-type control plants (Figure 4.2A). Thus, changes in cellular NAD(P)$^+$/NAD(P)H

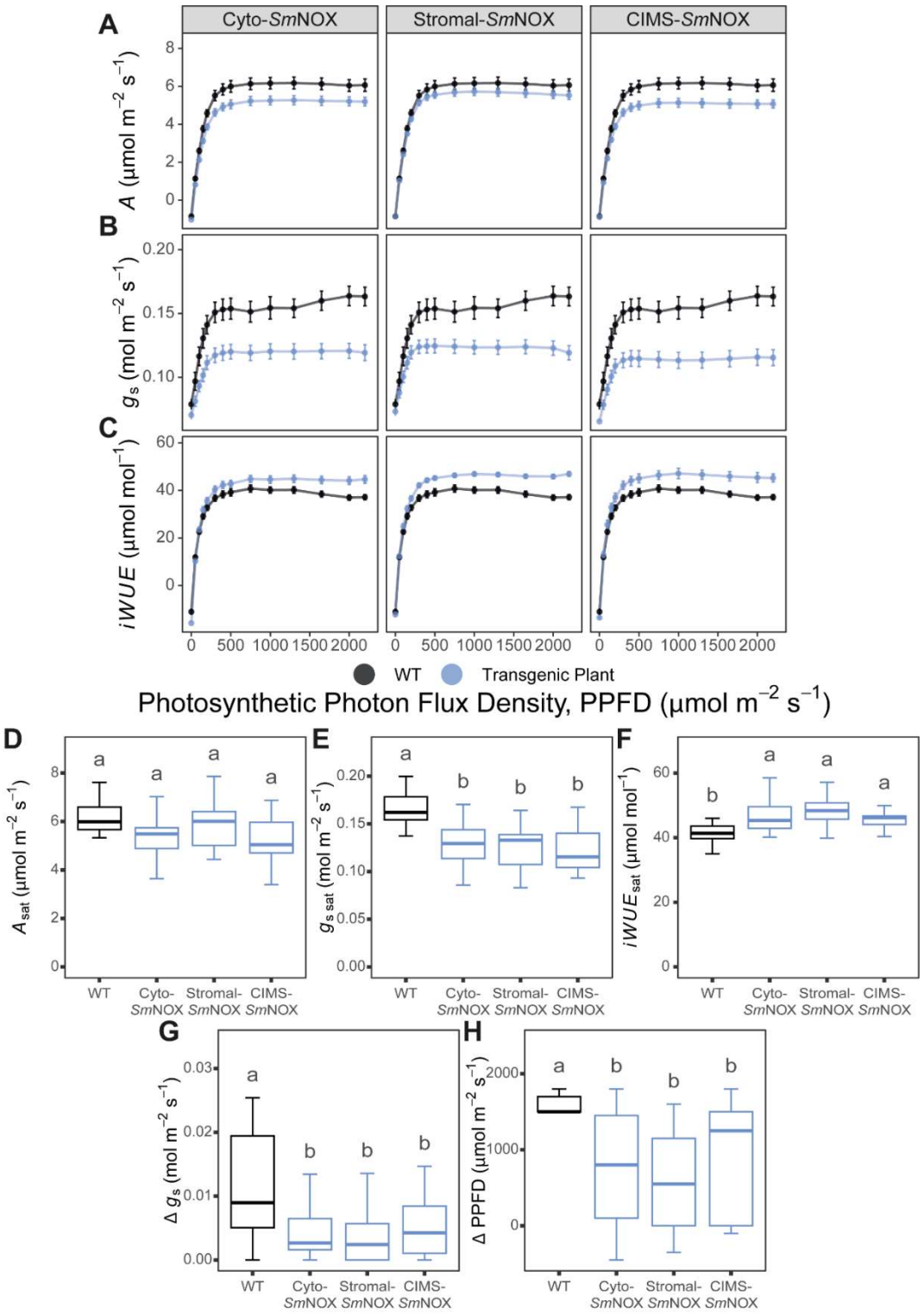


Figure 4.3. The response of plant gas exchange to light. A) CO_2 assimilation rate (A, µmol m^{-2} s^{-1}) at different PPFD in wild-type (black) and transgenic plants (blue). WT: wild-type. Cyto-*Sm*NOX: cytosolic *Sm*NOX. Stromal-*Sm*NOX: chloroplast stromal *Sm*NOX. CIMS-*Sm*NOX: chloroplast intermembrane space *Sm*NOX. Data represent mean ± 1 S.E. B) as in (A), but for stomatal conductance (g_s, mol m^{-2} s^{-1}). C) as in (A), but for intrinsic water-use efficiency (*iWUE*, µmol mol^{-1}). D) Boxplot depicting the light-saturated assimilation rate (A_{sat}, µmol m^{-2} s^{-1}). The colour scheme and abbreviations follow that described in (A). E) As in (D), but for the light-saturated stomatal

conductance ($g_{s\ sat}$, mol m^{-2} s^{-1}). F) As in (D), but for the light-saturated intrinsic water-use efficiency ($iWUE_{sat}$, µmol mol^{-1}). G) Boxplot depicting the difference in stomatal conductance between the light intensity at which 95% maximum assimilation rate and the light-saturated stomatal conductance are achieved (Δ g_s, mol m^{-2} s^{-1}). H) As in (G), but for the difference in light intensity at which 95% maximum assimilation rate and the light-saturated stomatal conductance are achieved (Δ PPFD, µmol m^{-2} s^{-1}). All data represent an average across three independent single copy lines (n = 7 – 8 plants line). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). The raw data can be found in ***Supplemental File 4 online***. **[*End of Figure Legend*].**

---

ratios, irrespective of the location of perturbation of that ratio, caused a suppression in stomatal development resulting in a reduced density of stomata on leaves.

Stomatal density changes have previously been linked with secondary changes in mesophyll development (such that leaves with reduced stomatal numbers are generally thinner and *vice versa* (Dow, Berry and Bergmann, 2017; Pérez-Bueno *et al.*, 2022)) which can have important consequences for plant physiology (Serna and Fenoll, 2000; Bergmann, 2004; Savaldi-Goldstein, Peto and Chory, 2007; Kondo *et al.*, 2010; Vile *et al.*, 2012; Lundgren *et al.*, 2019; Baillie and Fleming, 2020). Thus, we also investigated whether *Sm*NOX-expression was associated with an alteration in leaf width or cellular density by measuring leaf mass per area (LMA = leaf dry mass / leaf area). This revealed *Sm*NOX-expressing lines exhibited a small but significant reduction in LMA compared to wild-type (Figure 4.2B). Thus, expression of *Sm*NOX results in both a decrease in stomatal patterning and anticipated secondary alterations in leaf architecture.

4.3.5 SmNOX-induced changes in stomatal density improve plant photosynthetic water-use efficiency

As stomatal density is a key determinant of leaf gaseous conductance, we next characterised the effect of *Sm*NOX expression on plant physiological performance. To test this, we first assessed the photosynthetic parameters of *Sm*NOX-expressing lines along a light intensity gradient using an infrared open-gas exchange system (Figure 4.3A – 4.3C). This analysis revealed that there was a small reduction in CO_2 assimilation rate (A) between *Sm*NOX-expressing lines and wild-type plants

(Figure 4.3A). However, stomatal conductance (g_s) was substantially reduced across all *Sm*NOX-expressing lines (Figure 4.3B), consistent with these mutants exhibiting lower ratios of intercellular CO_2 to ambient CO_2 (C_i/C_a) and reduced rates of transpiratory water loss under high light (Appendix D [and ***Supplemental File 1 online***], Figure S5A and S5B). *Sm*NOX-expressing lines also exhibited comparable light-saturated rates of CO_2 assimilation (A_{sat}, Figure 4.3D, Appendix D [and ***Supplemental File 1 online***], table S2), and substantially lower light-saturated rates of stomatal conductance ($g_{s\ sat}$, Figure 4.3E, Appendix D [and ***Supplemental File 1 online***], table S2). As a consequence of these combined effects on *A* and g_s, all *Sm*NOX-expressing lines exhibited a higher intrinsic water use efficiency (*iWUE* = A/g_s) across a broad range of light intensities (Figure 4.3C and 4.3F, Appendix D [and ***Supplemental File 1 online***], table S2).

In addition to an enhanced *iWUE*, *Sm*NOX expression also resulted in tighter coordination between rates of *A* and g_s. Specifically, all mutants displayed a reduced "overshooting" of g_s at light intensities above those where 95% maximum assimilation was achieved (beyond which neither CO_2 delivery nor light is considered limiting) (Appendix D [and ***Supplemental File 1 online***]; Figure 4.3G and 4.3H; Appendix D [and ***Supplemental File 1 online***], Figure S6A – S6F, table S3). This improved *iWUE* and tighter coupling between *A* and g_s was not associated with a change in any other measured photosynthetic parameter including light compensation point, mitochondrial respiration in the light, apparent quantum yield (i.e., the efficiency by which light can be converted into ATP and NADPH), or the light-saturated rate of electron transport (Appendix D [and ***Supplemental File 1 online***], Figure S5C – S5F, table S2). Moreover, there was no consistent difference in any derived photosynthetic parameter when CO_2 assimilation was measured as a function of sub-stomatal CO_2 concentration (Appendix D [and ***Supplemental File 1 online***], Figure S7A – S7F, table S4). This included no change in the maximal CO_2 saturated assimilation rate, CO_2-compensation point, the transition point between rubisco-limited and RuBP regeneration-limited rates of photosynthesis, the rubisco Michaelis constant for CO_2 in the presence of 21% O_2 air, the carboxylation rate of rubisco, or the maximum electron transport rate through PSII (Appendix D [and ***Supplemental File 1 online***], Figure S7A – S7F, table S4). As anticipated, NOX activity was able to alter stomatal density, but was not sufficient to exert any measurable impact on mesophyll cell photosynthesis. Accordingly,

*Sm*NOX-mediated reduction in stomatal density drives both an enhanced photosynthetic water-use efficiency and a tighter coordination between rates of CO_2 assimilation and gas exchange.

4.3.6 SmNOX does not affect the kinetics of stomatal movement

We next sought to examine whether changes in stomatal density caused by *Sm*NOX expression were also associated with off-target impacts on stomatal behaviour. Specifically, we chose to assess the speed and magnitude of stomatal responses to light, given that this is one of the most important fluctuating variables that plants face under real-world conditions in the field. Light was also chosen because $NADP^+$ is the terminal electron acceptor in the photosynthetic electron transport chain and thus the impact of light on stomatal kinetics had potential to be perturbed by *Sm*NOX expression. For the purpose of this analysis, the experimental design of (McAusland *et al.*, 2016) was followed. In brief, plants were first acclimated at 100 µmol m^{-2} s^{-1} of light for 60 minutes, followed by a step increase in light intensity to 1000 µmol m^{-2} s^{-1} for 60 minutes, and a final step decrease in light intensity to 100 µmol m^{-2} s^{-1} for 30 minutes, with gas-exchange parameters recorded at one-minute intervals over this entire period.

This analysis revealed that all plants experience a normal sigmoidal response of g_s in response to step changes in irradiance (Figure 4.4A and 4.4B, Appendix D [and ***Supplemental File 1 online***], Figure S8). Moreover, subsequent model-based interrogation (Vialet-Chabrand, Dreyer and Brendel, 2013) of these data demonstrated that neither the amplitude nor the speed of stomatal responses was different between wild-type and transgenic plants. Specifically, no difference was found in either the modelled initial lag time in g_s, the maximum rate of change of g_s (calculated from the slope of the exponential phase of the curve), or the time taken for g_s to reach new steady-state upon the initial step increase in light intensity from 100 to 1000 µmol m^{-2} s^{-1} (Figure 4.4C – 4.4E, Appendix D [and ***Supplemental File 1 online***], table S5). Similarly, no difference could be detected in the time taken to reach new steady-state g_s upon the subsequent step decrease in light intensity from 1000 to 100 µmol m^{-2} s^{-1} (Figure 4.4F, Appendix D [and ***Supplemental File 1 online***], table S5). Thus, consistent with the lack of effect of *Sm*NOX on plant photosynthesis, *Sm*NOX expression did not impair the dynamic behaviour of stomatal movement.

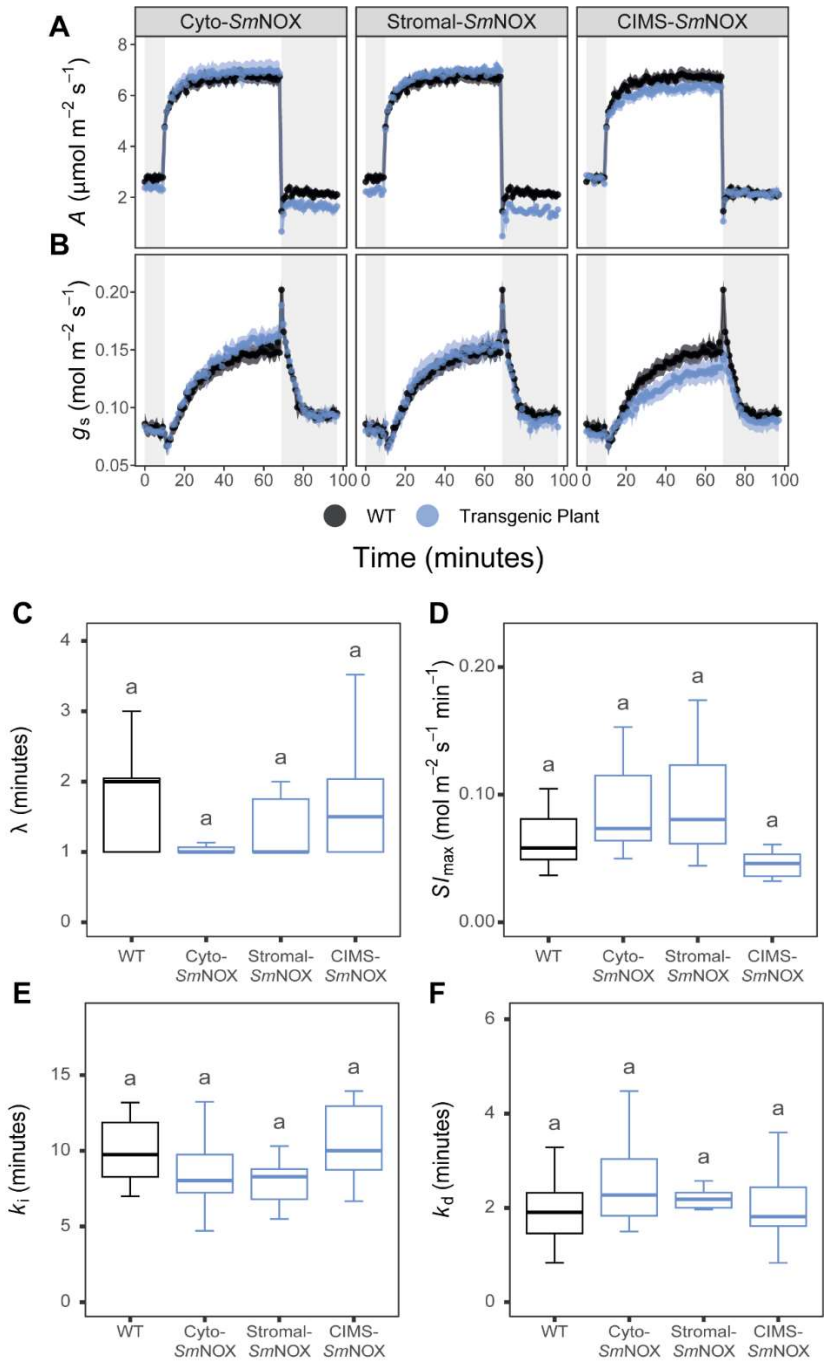


Figure 4.4. The temporal response of plant gas exchange to step changes in light intensity. A) The response of CO_2 assimilation rate (A, μmol m^{-2} s^{-1}) to step changes in light intensity in wild-type (black) and transgenic plants (blue). Shaded and unshaded areas represent periods of light intensity of 100 μmol m^{-2} s^{-1} and 1000 μmol m^{-2} s^{-1}, respectively. WT: wild-type. Cyto-*Sm*NOX: cytosolic *Sm*NOX. Stromal-*Sm*NOX: chloroplast stromal *Sm*NOX. CIMS-*Sm*NOX: chloroplast intermembrane space *Sm*NOX. Data represent mean ± 1 S.E. B) as in (A), but for stomatal conductance (g_s, mol m^{-2} s^{-1}). C) Boxplot depicting the estimated initial lag in response time of g_s (λ, minutes) after the step increase in light intensity from 100 μmol m^{-2} s^{-1} to 1000 μmol m^{-2} s^{-1}. The colour scheme and abbreviations follow that described in (A). D) As in (C), but for the estimated maximal rate of stomatal opening (SI_{max}, mol m^{-2} s^{-1} min^{-1}) after the step increase in light intensity from 100 μmol m^{-2} s^{-1} to 1000 μmol m^{-2} s^{-1}. E) As in (C), but for the estimated time taken to achieve new steady state stomatal

conductance (k_i, minutes) after the step increase in light intensity from 100 µmol m^{-2} s^{-1} to 1000 µmol m^{-2} s^{-1}. F) As in (C), but for the estimated time taken to achieve new steady state stomatal conductance (k_d, minutes) after the step decrease in light intensity from 1000 µmol m^{-2} s^{-1} to 100 µmol m^{-2} s^{-1}. All data represent an average across three independent single copy lines (n = 3 – 4 plants per line). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). The raw data can be found in ***Supplemental File 4 online.*** **[*End of Figure Legend*].**

4.3.7 SmNOX does not alter plant susceptibility to supra-optimal temperatures

One important function of the plant transpiration stream is to mediate the temperature control of plant aerial tissues through evaporative cooling. As such, plants exhibiting reduced stomatal numbers might be more susceptible to thermal damage of photosynthesis under supra-optimal temperatures, owing to impairment of their temperature control mechanism (Appendix D [and ***Supplemental File 1 online***], Figure S5B). To investigate this, we measured plant photosynthesis across step-wise increases in air temperatures from 20 to 45 °C. Under these conditions no difference could be detected in any photosynthesis parameter at any temperature between *Sm*NOX-expressing lines and wild type plants. In all plants, the rate of *A* increased with temperature from 20 °C, reaching an optimum around 30 °C, and then rapidly declined with increasing temperatures beyond this optimum consistent with thermal-induced inactivation of the photosynthetic machinery (Figure 4.5A). Despite an attenuated rate of g_s in mutant plants across all conditions, a comparable trend was also observed in this parameter between both mutant and wild-type plants (Figure 4.5B). Moreover, no difference could be detected in the measured leaf temperature of *Sm*NOX-expressing lines compared to wild-type plants at any given air temperature (Figure 4.5C). Thus, reductions in stomatal density in *Sm*NOX-expressing lines was not associated with either a change in leaf temperature or any change in the susceptibility to thermal damage of the photosynthetic machinery under supra-optimal temperatures.

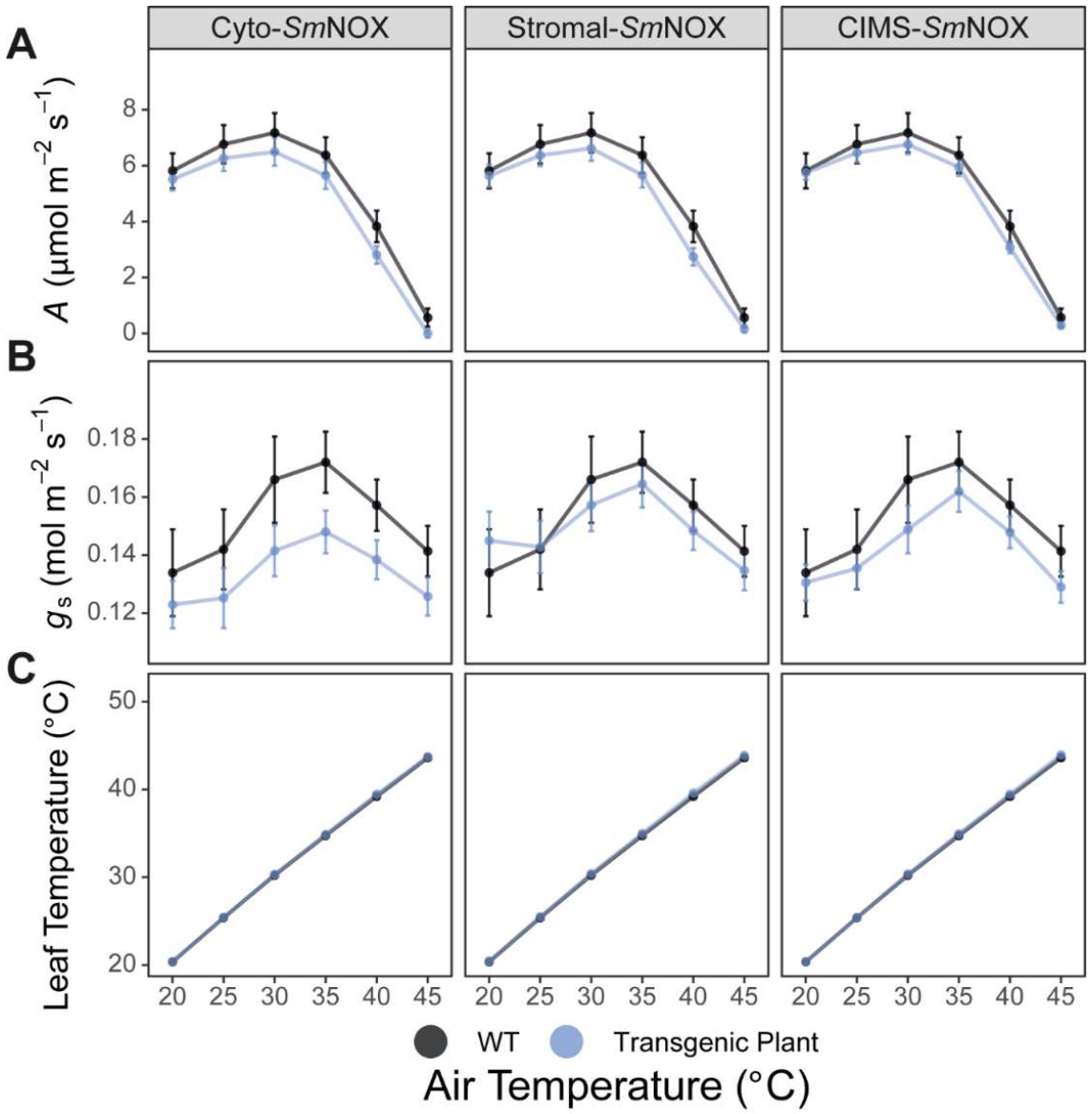


Figure 4.5. The response of plant gas exchange to air temperature (°C). A) CO_2 assimilation rate (*A*, µmol m^{-2} s^{-1}) at different air temperatures in wild-type (black) and transgenic plants (blue). WT: wild-type. Cyto-*Sm*NOX: cytosolic *Sm*NOX. Stromal-*Sm*NOX: chloroplast stromal *Sm*NOX. CIMS-*Sm*NOX: chloroplast intermembrane space *Sm*NOX. Data represent mean ± 1 S.E. B) as in (A), but for stomatal conductance (g_s, mol m^{-2} s^{-1}). C) as in (A), but for leaf temperature (°C). All data represent an average across three independent single copy lines (*n* = 3 – 5 plants per line). The raw data can be found in ***Supplemental File 4 online***.

4.3.8 SmNOX improves vegetative growth and biomass accumulation under water-deficit conditions

Based on the findings that *Sm*NOX expression caused an enhancement in *iWUE* without any off-target effects on photosynthetic efficiency, stomatal behaviour, or thermal sensitivity, we next sought to determine whether an impact on plant growth could also be observed. For this purpose, we grew plants in a fully randomised block design under long-day conditions with a 16-hour photoperiods and 250 µmol m^{-2} s^{-1} non-limiting light. Under these conditions, no difference in the vegetative growth or performance between transgenic *Sm*NOX-expressing plants and wild-type plants was detected under well-watered conditions (Appendix D [and ***Supplemental File 1 online***], Figure S9A – S9C). Given the higher *iWUE* of mutant plants, we next determined whether these improve photosynthesis-water relations resulted in any change in plant growth under water-deficit conditions. For this purpose, plant trays were watered prior to stratification, but water was then withheld for 15-days after the termination of stratification to maintain sub-optimal and deteriorating soil-water saturation levels during vegetative growth (after which watering was resumed when the top layer of soil was visibly dry every ~7 days). This assay revealed that *Sm*NOX expressing plants produced more biomass and larger rosettes compared to wild-type plants over equivalent periods of time (Figure 4.6A – 4.6E). This result was caused by *Sm*NOX expressing-lines maintaining higher relative growth rates during water-deficit (Figure 4.6F). Consistent with this enhancement in vegetative growth and tolerance to adverse water conditions, mutants also exhibited earlier times of bolting (Figure 4.6G) and flowering (Figure 4.6H), and produced visibly taller inflorescences relative to wild-type (Figure 4.6I). However, these alterations in vegetative growth and biomass did not translate to improved yields, and all *Sm*NOX expressing-lines exhibited a reduction in total seed mass (Figure 4.6J). Thus, *Sm*NOX-expressing plants exhibited enhanced biomass production and tolerance to water-deficit conditions but an overall impairment of seed production.

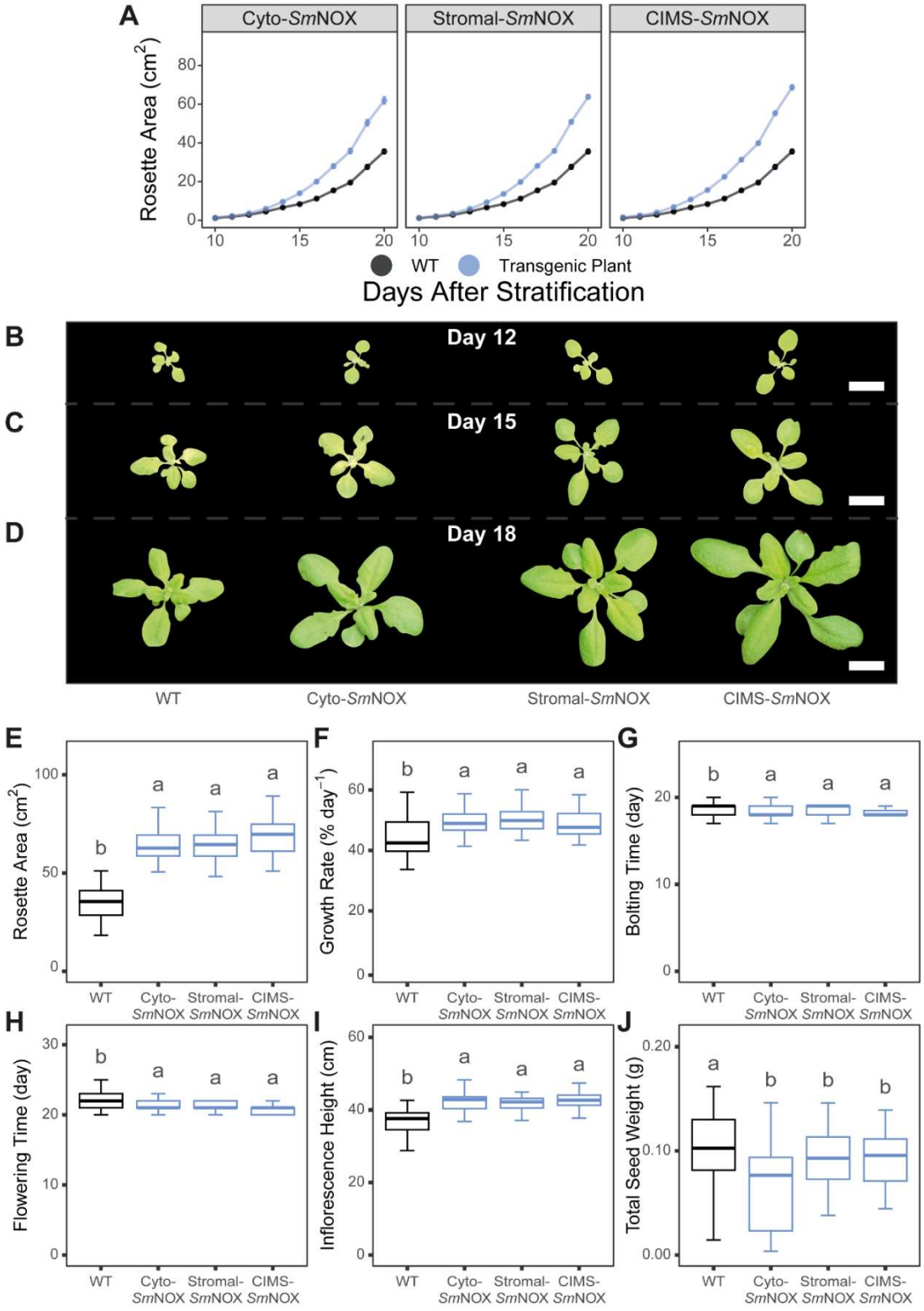


Figure 4.6. The vegetative and reproductive performance of plants under water-deficit conditions. (A) The increase in visible rosette area (cm^2) over time in wild-type (black) and transgenic plants (blue). WT: wild-type. Cyto-*Sm*NOX: cytosolic *Sm*NOX. Stromal-*Sm*NOX: chloroplast stromal *Sm*NOX. CIMS-*Sm*NOX: chloroplast intermembrane space *Sm*NOX. Data represent mean ± 1 S.E. B) Representative images of wild-type and transgenic plants taken on day 12 post stratification.

Scale bar = 2.5 cm. C) As in (B), but for day 15 post stratification. D) As in (B), but for day 18 post stratification. In parts B – D, the same individual plant is shown across the growing period and all images were generated by computationally removing background from the aerial photographs analysed for growth analyses. E) Boxplot depicting the maximum visible rosette area at the end of vegetative growth (day = 20). The colour scheme and abbreviations follow that described in (A). F) As in (E), but for the average percentage growth rate of plants computed over the vegetative growth phase (% increase in rosette area day^{-1}). G) As in (E), but for the time taken for bolting to occur in plants (days post stratification). H) As in (E), but for the time taken for flowering to occur in plants (days post stratification). I) As in (E), but for the maximum inflorescence height (cm) measured at day 32 post stratification. J) As in (E), but for the total seed weight harvested from plants after drying (g). All data represent an average across three independent single copy lines (n = 17 – 20 plants per line for all measurements, except total seed weight where n = 11 – 13 plants per line). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). The raw data can be found in ***Supplemental File 4 online***. **[*End of Figure Legend*].**

4.4 Discussion

Manipulation of stomatal density is a key target for enhancing crop productivity and climate resilience under current and future conditions (Bertolino, Caine and Gray, 2019; Leakey *et al.*, 2019). In the present study, we demonstrate that driving the expression of a water-forming NAD(P)H oxidase from *Streptococcus mutans* (*Sm*NOX) in stomatal precursor cells causes a reduction in stomatal density in *Arabidopsis thaliana*. We show that this reduced stomatal density phenotype occurs irrespective of whether *Sm*NOX is targeted to the cytosol, chloroplast stroma, or chloroplast intermembrane space. We also find that neither *Sm*NOX expression nor *Sm*NOX-mediated reductions in stomatal density had any detectable impact on plant CO_2 assimilation rate, the efficiency or thermal sensitivity of photosynthesis, the kinetics of stomatal movement, or plant growth under well-watered conditions. Nonetheless, we discover that all *Sm*NOX-expressing lines exhibit an enhanced intrinsic water-use efficiency. We also show that this improvement in water-use efficiency translates to an increase in plant vegetative growth and biomass accumulation under water-deficit conditions. Collectively, we

thus provide a novel metabolic engineering approach which can be leveraged to alter stomatal patterning and plant photosynthesis-water relations.

Nicotinamide adenine dinucleotide [NAD(H)] and nicotinamide adenine dinucleotide phosphate [NADP(H)] are ubiquitous coenzymes that facilitate cellular energy transfer, redox homeostasis, and signalling in living organisms (Smith *et al.*, 2021). Alongside these diverse roles, NAD^+ has recently been implicated as a negative regulator of stomatal development. Evidence in support of this hypothesis has arisen because perturbations in cellular redox status associated with translational repression of mitochondrial (*At*NDT1 and *At*NDT2) and peroxisomal (*At*PXN) NAD^+ transporter proteins, as well as poly (ADP-ribose) polymerase (PARP) proteins (PARP1, PARP2 and PARP3) involved in NAD^+ recycling, have all produced *A. thaliana* mutant plants with reduced stomatal densities (de Souza Chaves *et al.*, 2019; Feitosa-Araujo, da Fonseca-Pereira, *et al.*, 2020). The mechanism of this developmental effect is proposed to occur via abscisic acid (ABA) which is known to inhibit entry into the stomatal lineage by suppressing the activity of SPEECHLESS (SPCH) (Yoko Tanaka *et al.*, 2013), possibly via induction of the cyclin-dependent kinase (CDK) inhibitor gene ICK1 (Wang *et al.*, 1998; Chater *et al.*, 2014). This interaction between NAD^+ and ABA was postulated because NAD^+ transport and metabolism mutants were associated with both transcriptional changes in ABA metabolic genes and reduced SPCH expression (Feitosa-Araujo, da Fonseca-Pereira, *et al.*, 2020). Importantly, exogenous NAD^+ applications were also sufficient to drive a reduction in stomatal patterning in wild-type *Arabidopsis*, but not in mutants impaired in ABA biosynthesis or signalling (Feitosa-Araujo, da Fonseca-Pereira, *et al.*, 2020). In the present study, we build on this foundational biology by demonstrating that expression of a water-forming NAD(P)H oxidase in stomatal precursor cells is capable of driving a reduction in stomatal density through perturbation of endogenous cellular redox status. We further show that such a reduction in stomatal density is achieved irrespective of the subcellular compartment in which this enzyme is targeted. Thus, these results collectively demonstrate that NAD^+ is a viable metabolic handle which can be turned to alter stomatal patterning in synthetic biology applications.

The density of stomata imposes an upper anatomical constraint on leaf diffusive conductance and thus has important physiological implications for plant productivity (Haworth, Elliott-Kingston and

McElwain, 2011). Despite reduced stomatal density mutants having been generated in a diverse number of disparate species and studies (Bergmann, Lukowitz and Somerville, 2004; Bussis *et al.*, 2006; Pillitteri *et al.*, 2007; Hara *et al.*, 2007, 2009; Yoo *et al.*, 2010; Doheny-Adams *et al.*, 2012; Gudesblat *et al.*, 2012; Hepworth *et al.*, 2015; Franks *et al.*, 2015; Wang *et al.*, 2016; Hughes *et al.*, 2017; Yin *et al.*, 2017; Jalakas *et al.*, 2018; Mohammed *et al.*, 2019; Caine *et al.*, 2019; de Souza Chaves *et al.*, 2019; Dunn *et al.*, 2019; Feitosa-Araujo, da Fonseca-Pereira, *et al.*, 2020; Zhao *et al.*, 2024), the effects of this developmental alteration on plant photosynthesis, water-use efficiency, growth, and yield traits have proven variable across different studies and plant species (Lawson and Blatt, 2014). In the present study, we characterise the gas exchange properties of our reduced stomatal density mutants across a range of environmental conditions to shed light on their underlying physiology. We demonstrate that these reduced stomatal density plants exhibit an increased intrinsic water-use efficiency (*iWUE*) by maintaining wild-type rates of CO_2 assimilation at lower respective stomatal gaseous conductance. Reductions in stomatal density also caused a tighter co-ordination between the rates of CO_2 assimilation and stomatal conductance, such that a reduced extent of water loss occurred after light-saturated rates of CO_2 assimilation were achieved. However, decreasing stomatal density did not result in any change in the efficiency or thermal sensitivity of photosynthesis and CO_2 assimilation, or the kinetics of stomatal movements. Together, these results align with recent findings from Epidermal Patterning Factor mutants which similarly report reductions in stomatal density without any impact on plant CO_2 assimilation (Franks *et al.*, 2015; Hughes *et al.*, 2017; Caine *et al.*, 2019; Harrison *et al.*, 2020). The biological basis of these collective observations is intriguing. As the rate of photosynthesis at any one time is constrained by the most limiting factor of chloroplast CO_2 delivery (including both stomatal conductance and mesophyll conductance components) and photosynthetic biochemistry (Gago *et al.*, 2019), this phenomenon could be explained if stomata are either not limiting to plant photosynthesis under the study conditions or if secondary adjustments in leaf architecture compensate for changes in stomatal density (Harrison *et al.*, 2020). Future research efforts will be required to further elucidate the complex interplay between stomata, leaf architecture and plant physiology.

In addition to the lack of effect of reduced stomatal density on leaf photosynthesis, CO_2 assimilation was also not impacted by the expression of SmNOX under the control of the CAB3 promoter. Thus, whilst *Sm*NOX was able to perturb NAD^+ levels to mediate a reduction in stomatal density, it was not sufficient to impair photosynthesis. The reason for this is likely that the large pool sizes of NADH (24 – 55 µM) and NADPH (140 – 180 µM) (Igamberdiev and Gardeström, 2003) and the extraordinary high rates of photosynthetic NADPH production (~70 µmol g^{-1} s^{-1}) (Wieloch and Sharkey, 2022) would together mask the effect of *Sm*NOX activity in the mesophyll tissue. By contrast, epidermal cells contribute minimally toward total plant photosynthesis and contain small chloroplasts with high stroma to grana ratios which would produce much lower rates of NADPH (Barton *et al.*, 2016, 2018). Thus, although *Sm*NOX is expressed to a high level throughout diverse photosynthesis tissues using the CAB3 promoter, the impact of this enzyme on cellular NAD^+ status would be largest in stomatal progenitor cells.

Consistent with the lack of impairment of stomatal density or *Sm*NOX expression on photosynthetic CO_2 assimilation, transgenic plants did not exhibit any difference in plant growth or biomass accumulation under well-watered conditions. However, increased vegetative growth and biomass accumulation was observed under water-deficit conditions. This improved tolerance to adverse water conditions is caused by the more conservative water-input requirements of these plants, driven by both the reduced rates of transpiration and improvement in photosynthetic water-use efficiency and is in agreement with other analyses of reduced stomatal density mutants (Wang *et al.*, 2016; Hughes *et al.*, 2017; Caine *et al.*, 2019; Mohammed *et al.*, 2019). Moreover, the reduced leaf mass per area measured in *Sm*NOX-expressing lines would also favour biomass accumulation as this requires a lower resource investment per unit leaf tissue produced. Despite this improved tolerance to adverse water conditions, all *Sm*NOX-expressing lines exhibited a lower overall seed yield. Similar impairments in plant reproductive traits were also reported in NAD^+ carrier mutants (de Souza Chaves *et al.*, 2019; Feitosa-Araujo, De Souza Chaves, *et al.*, 2020) and are likely caused by the role of NAD^+ in seed (Chai *et al.*, 2005; Katoh *et al.*, 2006) and pollen development (Cárdenas *et al.*, 2006; Hashida *et al.*, 2007, 2013; Hashida, Kawai-Yamada and Uchimiya, 2013). It would be

interesting to investigate in future work whether this yield impairment is unavoidable, or whether it could be instead mitigated by use of other promoters with reduced off-target expression effects.

As global food security is increasingly threatened by climate change, developing more climate-resilient crops presents an important goal for crop breeding and engineering efforts. Altering stomatal density represents a promising avenue to optimise the relationship between plant carbon assimilation, water demand, growth, and yield. Here, we provide a novel metabolic engineering strategy which can be exploited to reduce plant stomatal density by altering the endogenous NAD^+ status of stomatal precursor cells. We demonstrate that this approach produces plants with improved photosynthesis-water relations and improved productivity under water-deficit conditions without any apparent cost on photosynthetic carbon assimilation or growth under normal conditions. Thus, this approach reveals a potential novel way in which plants can be altered to enhance climate resilience.

4.5 Materials and Methods

4.5.1 Gene accession numbers

The following list comprises the accession numbers of the genes used in this study: *Sm*NOX2 (WP_002268749.1); *At*RBCS1A (At1g67090); *At*PSBQA (At4g21280)*; At*OEP9 (At1g16000).

4.5.1 Gene synthesis

A variant of the NAD(P)H oxidase (NOX) enzyme encoded by the Streptococcus mutans *nox2* gene but containing amino acid substitutions D192A/V193R/V194H/A199R was chosen for use in this study (Petschacher *et al.*, 2014). This gene (hereinafter, termed *Sm*NOX) was codon optimised for expression in *Arabidopsis thaliana* and domesticated to remove the type IIS restriction enzymes BsaI or BpiI. Appropriate flanking regions compatible with Golden Gate Assembly (Marillonnet and Engler, 2014) were added to the 5' and 3' regions, and the resulting coding sequence was commercially synthesized (Marillonnet and Engler, 2014) (TWIST Bioscience, USA). The full-length coding sequence of *AtOEP9* and *At*RBCS1A, as well as a chloroplast stromal transit peptide from *AtPSBQA* (containing the first 87 amino acids of *AtPSBQA* including the intact stromal import domain but lacking a complete thylakoid targeting motif (Ma *et al.*, 2021)) were subject to these same steps for codon optimisation, domestication, addition of compatible flanking regions for Golden Gate assembly

(Marillonnet and Engler, 2014), and gene synthesis. The full-length sequences of each these genes are found in ***Supplemental File 2 online***.

4.5.3 Construct design and assembly

Genetic constructs for stable expression in plants were generated by cloning the coding sequence of *Sm*NOX downstream of the *A. thaliana* chlorophyll a/b-binding protein 3 (*At*CAB3; pTEI071) and upstream of the *A. tumefaciens* nopaline synthase terminator (nosT; piCH41421) in the Level 1 backbone position forward 2 vector pICH47742. Compositional forms of this Level 1 module were also generated in which *Sm*NOX was translationally fused at the N-terminus to either the *At*PSBQA transit peptide or the *At*OEP9 full-length coding sequence, respectively. Each of these constructs were subsequently assembled into Level 2 Golden Gate modules using the plant transformation vector acceptor backbone pAGM37443 (Grützner *et al.*, 2021) with the pFAST-R seed-coat specific fluorescent selection cassette (Engler *et al.*, 2014) (pICSL70008) and the end Linker position 3 insert (pICH49266).

Genetic constructs for transient expression in protoplasts were generated by cloning the coding sequence of *Sm*NOX translationally fused to a C-terminal enhanced green fluorescent protein (GFP; pJOG176) downstream of the cauliflower mosaic virus 35S promoter (CaMV35S; EC15058) and upstream of the *A. tumefaciens* nosT (piCH41421) in the Level 1 backbone position forward 2 vector pICH47742. As above, compositional forms of this were also generated in which *Sm*NOX was translationally fused at the N-terminus to the *At*PSBQA transit peptide or the *At*OEP9 full-length coding sequence, respectively. Positive control constructs were also cloned using the same promoter, terminator, and backbone vector but containing only the coding sequence of the GFP (cytosolic fluoresce control) or the coding sequence of GFP translationally fused at the N-terminus to the full-length *At*RBCS1A gene sequence (chloroplast stroma fluoresce control). All cloning reactions were performed following the Golden Gate one-step one-pot protocol (Engler *et al.*, 2009) as described in Appendix D [and ***Supplemental File 1 online***]. The provenance of commercial vectors used during cloning are also provided in Appendix D [and ***Supplemental File 1 online***]. Full-length sequences of each expression cassette generated is provided in ***Supplemental File 2 online***. The full set of oligonucleotides used for cloning, colony PCR, and Sanger sequencing are found in

Supplemental File 3 online. The thermocycler conditions used for all PCR reactions are also described in ***Supplemental File 3 online***.

4.5.4 Protoplast isolation, transformation, and visualisation

Mesophyll protoplasts were isolated from leaves of 4-week old wild-type *Arabidopsis thaliana* Col-0 plants following the 'Tape-*Arabidopsis* Sandwich' method (Wu *et al.*, 2009). Protoplast transformation was performed using the calcium/polyethylene glycol-mediated method (Yoo, Cho and Sheen, 2007), as modified in (Wu *et al.*, 2009). Transformed protoplasts were incubated in the dark overnight at 25°C prior to imaging. Cell fluorescence was visualised using a Zeiss LSM 880 Airy Scan confocal microscope. Zeiss ZEN software (Zeiss) was used for confocal laser scanning microscopy and subsequent image processing.

4.5.5 Plant material and stable transformation

Stable transformation of *A. thaliana* ecotype Columbia (Col-0) background was performed using the agrobacterium-mediated floral dip method (Clough and Bent, 1998). Positive transformants were initially screened based upon seed fluorescence visualised by epifluorescence microscopy (Leica M165 stereo microscope), and confirmed by genotyping PCR and gel electrophoresis using the OneTaq® 2X Master Mix with Standard Buffer (NEB, Catalog #M0482S) following manufacturer's instructions for reaction set-up and the oligonucleotides and thermocycler conditions described in ***Supplemental File 3 online***. Single insertion events were subsequently identified by TaqMan® real-time PCR with fluorescent probes and primers which bind to the TagRFP selectable marker (Idna GENETICS) and homozygous individuals were isolated. For the above purposes of genotyping PCR and TaqMan® real-time PCR, genomic DNA was isolated from mature leaves of T1 plants using the cetyl trimethylammonium bromide (CTAB) method (Murray and Thompson, 1980). In total, three single-copy homozygous lines were taken forward as the basis of the analysis herein for each *Sm*NOX-expressing transgenic plant (cyto-*Sm*NOX, stromal-*Sm*NOX and CIMS-*Sm*NOX). All experiments described are performed on T3 and T4 generation plants.

4.5.6 Stomatal density measurements

Epidermal impressions from the abaxial surface of healthy and fully-expanded excised leaves of 8-week old plants were obtained following the clear nail polish method (Wu and Zhao, 2017). All leaves

used for this analysis were of similar size. After mounting on microscope slides, stomatal density was measured using a Leica DMR8 light microscope at ×20 magnification under brightfield view. For each sample, stomatal density counts were calculated across five independent fields of view of the same leaf and across 9 – 10 biological replicates per line.

4.5.7 Leaf mass area measurements

A leaf puncher was used to harvest a fixed area of tissue from healthy and fully-expanded leaves of 8-week-old plants. Leaf dry mass was subsequently determined after 48 hours incubation at 65 °C. All leaves were of similar size and were harvested in the middle of the photoperiod. Leaf mass area (LMA, $g\,m^{-2}$) was subsequently determined as the fixed leaf area divided by leaf dry mass.

4.5.8 In vitro enzyme assays

A consistent area of leaf tissue was harvested from healthy and fully-expanded leaves of 6-week-old plants using a leaf puncher during the middle of the photoperiod. Leaf material was flash frozen in liquid nitrogen, ground using metal beads and a tissue lyser (Thermo Scientific, TissueLyser II), and resuspended in 500 µL ice cold extraction buffer (50 mM potassium phosphate, 5 mM MgCl2, 1 mM EDTA, 3% Triton X-100) with 1x protease inhibitor cocktail (Sigma, Catalog #P9599-5ML). Homogenate was centrifuged for 1 minute at 12,000 rpm at 4°C, and total crude protein extract was isolated from supernatant.

Enzyme assays were performed following the method described in (Petschacher *et al.*, 2014) with some alterations. In brief, all reactions were performed using a total volume of 200 µL in unsealed 96 well plate at 25°C. Typical reaction mixtures contained 50 mM potassium phosphate buffer (pH 7.5) and 250 µM of either NADH or NADPH, were vortexed for saturation with oxygen in air (250 µM), and were initiated by addition of 30 µL undiluted crude leaf lysate. Negative controls containing ddH_2O in the place of leaf lysate were included in every assay. Consumption of NADH or NADPH was monitored spectrophotometrically at 340 nm using a plate reader (BMG Labtech FLUOstar Omega). Each biological replicate was repeated over two technical replicates wells and rates of enzyme reaction were calculated over the linear part of the curve over a duration of at least 5 minutes. Reaction rates were normalized per unit NADH/NADPH oxidation by standard curve.

4.5.9 Leaf gas exchange

Gas exchange analyses were performed using a LI-6800 open-path gas exchange system equipped with a 6800-01A fluorometer head (LI-COR, USA) using the 2 cm^2 leaf chamber and 90% red and 10% blue actinic light under normal 21% O_2 air. The youngest fully-expanded leaf was used for each plant and all measurements were taken prior to onset of flowering. Downstream statistical analysis, model fitting, and parameter estimation from leaf gas exchange data was performed as described in Appendix D [and ***Supplemental File 1 online***].

Response of plant gas exchange to light followed the method of (López-Calcagno *et al.*, 2020) with some alterations. In brief, plants were first adapted at 2,200 µmol m^{-2} s^{-1} for 20 minutes, and photosynthesis was then measured at step-wise intervals of 2200, 2000, 1650, 1300, 1000, 750, 500, 400, 300, 200, 150, 100, 50, 0 and 2200 µmol m^{-2} s^{-1} light. The environment was held constant at 400 µmol mol^{-1} CO_2, 30°C leaf temperature, 55% relative air humidity, 500 µmol s^{-1} flow rate, and 10,000 rpm fan speed. No matching was performed during the analysis and a 4-minute wait time was used between each light condition.

For the response of plant gas exchange to intercellular CO_2, plants were first adapted at 400 µmol mol^{-1} CO_2 for 20 minutes and photosynthesis was then measured at step-wise intervals of 400, 0, 50, 75, 100, 200, 400, 800, 1200, 1600, 400 µmol mol^{-1} reference atmospheric CO_2 concentrations. The environment was held constant at 1500 µmol m^{-2} s^{-1} light, 30°C leaf temperature, 55% relative air humidity, 500 µmol s^{-1} flow rate, and 10,000 rpm fan speed. Matching of CO_2 and H_2O was performed before each measurement with a minimum and maximum wait time of 3 and 4 minutes, respectively.

Investigation of stomatal kinetics in response to step changes in light followed the method of (McAusland *et al.*, 2016). In brief, plants were first adapted to 100 µmol m^{-2} s^{-1} light for 60 minutes and were then subject to a step increase to 1000 µmol m^{-2} s^{-1} for 60 minutes followed by a step decrease back to 100 µmol m^{-2} s^{-1} for 30 minutes. The environment was held constant at 400 µmol mol^{-1} CO_2, 30°C leaf temperature, 1.8 kPa vapour pressure deficit, 500 µmol s^{-1} flow rate, and 10,000 rpm fan speed. No matching was performed during the analysis and measurements were taken every minute.

For the response of plant gas exchange to temperature, plants were first adapted at 20°C for 20 minutes and photosynthesis was then measured at step-wise intervals of 20, 25, 30, 35, 40 and 45°C reference air temperatures. The environment was held constant at 1500 µmol m^{-2} s^{-1} light, 400 µmol mol^{-1} CO_2, 500 µmol s^{-1} flow rate, and 10,000 rpm fan speed. As previously described in the thermal analysis of (Busch and Sage, 2017), it is difficult to maintain VPD at temperatures above 30°C. Thus, leaf vapour pressure deficit (VPD) was maintained at 1.5 kPa for measurements at 20, 25 and 30°C, 2.5 kPa for the measurement at 35°C, 3.5 kPa for the measurement at 40°C and 4.5 kPa for the measurement at 45 °C. Matching of CO_2 and H_2O was performed before each measurement with a minimum and maximum wait time of 10 and 14 minutes, respectively.

4.5.10 Plant growth conditions

Seeds were grown in individual pots directly on soil (Levington Seed modular compost) and were stratified for 2 days at 4°C to overcome dormancy. Unless explicitly stated, plants were grown in climate-controlled environment conditions in floor-standing growth chambers (Aralab FitoclimaD1200 PLH) under short-day conditions (8-hour light/16-hour dark cycle) at 21°C, ambient 400 ppm CO_2, 50% relative humidity, and 125 µmol m^{-2} s^{-1} light and were maintained under well-watered conditions.

4.5.11 Plant growth assays

For analysis of plant growth, after stratification plants were transferred to controlled environment chambers (CERs) at The University of Oxford and grown under a 16-hour/8-hour day/night period, 21°C, and 250 µmol m^{-2} s^{-1} light. To minimize confounding and positional effects, plant genotypes were grown in a fully randomised block design and the position of trays were rotated every 2 days. For well-watered experiments, plants were watered on a weekly basis as usual from the base of the tray. For water-deficit growth experiment, plant trays were watered prior to stratification as above, but water was then withheld completely for 15 days after germination to maintain sub-optimal and deteriorating soil-water saturation levels during vegetative growth. After this period, watering was resumed following the method described in the well-watered growth experiments.

Aerial photographs were taken daily during vegetative growth. Image processing was performed to correct for orientation and perspective using Yet Another Scanning Wizard (YASW)

(https://github.com/ImageProcessing-ElectronicPublications/yasw). Computational analysis was performed to derive growth parameters. Green pixels were isolated from images using an in-built script, and visible rosette area was computed with ImageJ software(Schneider, Rasband and Eliceiri, 2012) using the in-built 'Analyze Particles' function. Time of bolting was measured as the number of days after stratification when the inflorescence was greater than 1 cm in height above the rosette. Time of flowering time was measured as the number of days after stratification at which the first flower was open and petals were visible. Plant height was measured as the maximum height of the primary inflorescence from the base of the rosette at 32 days post stratification. Total seed weight was measured after allowing plants to completely dry for a month.

4.6 References

Acevedo-Siaca, L.G. et al. (2022) 'Guard-cell-targeted overexpression of Arabidopsis Hexokinase 1 can improve water use efficiency in field-grown tobacco plants', Journal of Experimental Botany, 73(16), pp. 5745–5757. Available at: https://doi.org/10.1093/jxb/erac218.

Adrian, J. et al. (2015) 'Transcriptome dynamics of the stomatal lineage: Birth, amplification, and termination of a self-renewing population', Developmental Cell, 33(1), pp. 107–118. Available at: https://doi.org/10.1016/j.devcel.2015.01.025.

Antunes, W.C. et al. (2012) 'Changes in stomatal function and water use efficiency in potato plants with altered sucrolytic activity', Plant, Cell and Environment, 35(4), pp. 747–759. Available at: https://doi.org/10.1111/j.1365-3040.2011.02448.x.

Antunes, W.C. et al. (2017) 'Guard cell-specific down-regulation of the sucrose transporter SUT1 leads to improved water use efficiency and reveals the interplay between carbohydrate metabolism and K+ accumulation in the regulation of stomatal opening', Environmental and Experimental Botany, 135, pp. 73–85. Available at: https://doi.org/10.1016/j.envexpbot.2016.12.004.

Baillie, A.L. and Fleming, A.J. (2020) 'The developmental relationship between stomata and mesophyll airspace', New Phytologist, 225(3), pp. 1120–1126. Available at: https://doi.org/10.1111/nph.16341.

Bakker, J.C. (1991) 'Effects of humidity on stomatal density and its relation to leaf conductance', Scientia Horticulturae, 48(3–4), pp. 205–212. Available at: https://doi.org/10.1016/0304-4238(91)90128-L.

Barton, K.A. et al. (2016) 'Epidermal pavement cells of arabidopsis have chloroplasts', Plant Physiology, 171(2), pp. 723–726. Available at: https://doi.org/10.1104/pp.16.00608.

Barton, K.A. et al. (2018) 'Chloroplast behaviour and interactions with other organelles in Arabidopsis thaliana pavement cells', Journal of Cell Science, 131(2). Available at: https://doi.org/10.1242/jcs.202275.

Beerling, D.J., McElwain, J.C. and Osborne, C.P. (1998) 'Stomatal responses of the "living fossil" Ginkgo biloba L. to changes in atmospheric CO2 concentrations', Journal of Experimental Botany, 49(326), pp. 1603–1607. Available at: https://doi.org/10.1093/jxb/49.326.1603.

Beerling, D.J., Osborne, C.P. and Chaloner, W.G. (2001) 'Evolution of leaf-form in land plants linked to atmospheric CO2 decline in the Late Palaeozoic era', Nature, 410(6826), pp. 352–354. Available at: https://doi.org/10.1038/35066546.

Bergmann, D.C. (2004) 'Integrating signals in stomatal development', Current Opinion in Plant Biology, 7(1), pp. 26–32. Available at: https://doi.org/10.1016/j.pbi.2003.10.001.

Bergmann, D.C., Lukowitz, W. and Somerville, C.R. (2004) 'Stomatal development and pattern controlled by a MAPKK kinase', Science, 304(5676), pp. 1494–1497. Available at: https://doi.org/10.1126/science.1096014.

Bertolino, L.T., Caine, R.S. and Gray, J.E. (2019) 'Impact of stomatal density and morphology on water-use efficiency in a changing world', Frontiers in Plant Science, 10. Available at: https://doi.org/10.3389/fpls.2019.00225.

de Boer, H.J. et al. (2016) 'Optimal allocation of leaf epidermal area for gas exchange', New Phytologist, 210(4), pp. 1219–1228. Available at: https://doi.org/10.1111/nph.13929.

Boxall, S.F. et al. (2020) 'Kalanchoë PPC1 is essential for crassulacean acid metabolism and the regulation of core circadian clock and guard cell signaling genes[CC-BY]', Plant Cell, 32(4), pp. 1136–1160. Available at: https://doi.org/10.1105/TPC.19.00481.

Brown, D.M., Upcroft, J.A. and Upcroft, P. (1996) 'A H2O-producing NADH oxidase from the protozoan parasite Giardia duodenalis', European Journal of Biochemistry, 241(1), pp. 155–161. Available at: https://doi.org/10.1111/j.1432-1033.1996.0155t.x.

Buckley, C.R., Caine, R.S. and Gray, J.E. (2020) 'Pores for Thought: Can Genetic Manipulation of Stomatal Density Protect Future Rice Yields?', Frontiers in Plant Science, 10. Available at: https://doi.org/10.3389/fpls.2019.01783.

Busch, F.A. and Sage, R.F. (2017) 'The sensitivity of photosynthesis to O2 and CO2 concentration identifies strong Rubisco control above the thermal optimum', New Phytologist, 213(3), pp. 1036–1051. Available at: https://doi.org/10.1111/nph.14258.

Bussis, D. et al. (2006) 'Stomatal aperture can compensate altered stomatal density in Arabidopsis thaliana at growth light conditions', Functional Plant Biology, 33, pp. 1037–1043.

Caine, R.S. et al. (2019) 'Rice with reduced stomatal density conserves water and has improved drought tolerance under future climate conditions', New Phytologist, 221(1), pp. 371–384. Available at: https://doi.org/10.1111/nph.15344.

Cárdenas, L. et al. (2006) 'NAD(P)H oscillates in pollen tubes and is correlated with tip growth', Plant Physiology, 142(4), pp. 1460–1468. Available at: https://doi.org/10.1104/pp.106.087882.

Carroll, S. et al. (2022) 'Altering arabinans increases Arabidopsis guard cell flexibility and stomatal opening', Current Biology, 32(14), pp. 3170-3179.e4. Available at: https://doi.org/10.1016/j.cub.2022.05.042.

Castillo-Villanueva, A. et al. (2016) 'Cloning, Expression and Characterization of Recombinant, NADH Oxidase from Giardia lamblia', Protein Journal, 35(1), pp. 24–33. Available at: https://doi.org/10.1007/s10930-015-9643-9.

Chai, M.F. et al. (2005) 'NADK2, an Arabidopsis chloroplastic NAD kinase, plays a vital role in both chlorophyll synthesis and chloroplast protection', Plant Molecular Biology, 59(4), pp. 553–564. Available at: https://doi.org/10.1007/s11103-005-6802-y.

Chater, C. et al. (2015) 'Elevated CO2-Induced Responses in Stomata Require ABA and ABA Signaling', Current Biology, 25(20), pp. 2709–2716. Available at: https://doi.org/10.1016/j.cub.2015.09.013.

Chater, C.C.C. et al. (2014) 'Putting the brakes on: Abscisic acid as a central environmental regulator of stomatal development', New Phytologist, 202(2), pp. 376–391. Available at: https://doi.org/10.1111/nph.12713.

Clark, J.W. et al. (2022) 'The origin and evolution of stomata', Current Biology, 32(11), pp. R539–R553. Available at: https://doi.org/10.1016/j.cub.2022.04.040.

Clough, S.J. and Bent, A.F. (1998) 'Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana', Plant Journal, 16(6), pp. 735–743. Available at: https://doi.org/10.1046/j.1365-313X.1998.00343.x.

Condon, A.G. et al. (2002) 'Improving Intrinsic Water-Use Efficiency and Crop Yield', Crop Science, 42(1), p. 122. Available at: https://doi.org/10.2135/cropsci2002.0122.

Cracan, V. et al. (2017) 'A genetically encoded tool for manipulation of NADP+/NADPH in living cells', Nature Chemical Biology, 13(10), pp. 1088–1095. Available at: https://doi.org/10.1038/nchembio.2454.

Daloso, D.M. et al. (2016) 'Guard cell-specific upregulation of sucrose synthase 3 reveals that the role of sucrose in stomatal function is primarily energetic', New Phytologist, 209(4), pp. 1470–1483. Available at: https://doi.org/10.1111/nph.13704.

Dishisha, T. et al. (2019) 'Lactobacillus reuteri NAD(P)H oxidase: Properties and coexpression with propanediol-utilization enzymes for enhancing 3-hydroxypropionic acid production from 3-hydroxypropionaldehyde', Journal of Biotechnology, 289, pp. 135–143. Available at: https://doi.org/10.1016/j.jbiotec.2018.11.010.

Dittberner, H. et al. (2018) 'Natural variation in stomata size contributes to the local adaptation of water-use efficiency in Arabidopsis thaliana', Molecular Ecology, 27(20), pp. 4052–4065. Available at: https://doi.org/10.1111/mec.14838.

Doheny-Adams, T. et al. (2012) 'Genetic manipulation of stomatal density influences stomatal size, plant growth and tolerance to restricted water supply across a growth carbon dioxide gradient', Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1588), pp. 547–555. Available at: https://doi.org/10.1098/rstb.2011.0272.

Dow, G.J., Berry, J.A. and Bergmann, D.C. (2014) 'The physiological importance of developmental mechanisms that enforce proper stomatal spacing in Arabidopsis thaliana', New Phytologist, 201(4), pp. 1205–1217. Available at: https://doi.org/10.1111/nph.12586.

Dow, G.J., Berry, J.A. and Bergmann, D.C. (2017) 'Disruption of stomatal lineage signaling or transcriptional regulators has differential effects on mesophyll development, but maintains coordination of gas exchange', New Phytologist, 216(1), pp. 69–75. Available at: https://doi.org/10.1111/NPH.14746.

Driesen, E. et al. (2020) 'Influence of environmental factors light, co2, temperature, and relative humidity on stomatal opening and development: A review', Agronomy, 10(12). Available at: https://doi.org/10.3390/agronomy10121975.

Dunn, J. et al. (2019) ‘Reduced stomatal density in bread wheat leads to increased water-use efficiency’, Journal of Experimental Botany, 70(18), pp. 4737–4747. Available at: https://doi.org/10.1093/jxb/erz248.

Edwards, D., Kerp, H. and Hass, H. (1998) ‘Stomata in early land plants: An anatomical and ecophysiological approach’, Journal of Experimental Botany, 49(SPEC. ISS.), pp. 255–278. Available at: https://doi.org/10.1093/jxb/49.special_issue.255.

Eiteman, M.A. et al. (2006) ‘Overflow Metabolism in Escherichia coli during Steady-State Growth : Transcriptional Regulation and Effect of the Redox Ratio’, Applied and environmental microbiology, 72(5), pp. 3653–3661.

Engler, C. et al. (2009) ‘Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes’, PLoS ONE, 4(5). Available at: https://doi.org/10.1371/journal.pone.0005553.

Engler, C. et al. (2014) ‘A Golden Gate modular cloning toolbox for plants’, ACS Synthetic Biology, 3(11), pp. 839–843. Available at: https://doi.org/10.1021/sb4001504.

Falquetto-Gomes, P. et al. (2024) ‘From epidermal cells to functional pores: Understanding stomatal development’, Journal of Plant Physiology, 292. Available at: https://doi.org/10.1016/j.jplph.2023.154163.

Fanourakis, D., Heuvelink, E. and Carvalho, S.M.P. (2013) ‘A comprehensive analysis of the physiological and anatomical components involved in higher water loss rates after leaf development at high humidity’, Journal of Plant Physiology, 170(10), pp. 890–898. Available at: https://doi.org/10.1016/j.jplph.2013.01.013.

Farquhar, G.D. and Richards, R.A. (1984) ‘Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes.’, Australian Journal of Plant Physiology, 11(6), pp. 539–552. Available at: https://doi.org/10.1071/PP9840539.

Feitosa-Araujo, E., da Fonseca-Pereira, P., et al. (2020) ‘Changes in intracellular NAD status affect stomatal development in an abscisic acid-dependent manner’, Plant Journal, 104(5), pp. 1149–1168. Available at: https://doi.org/10.1111/tpj.15000.

Feitosa-Araujo, E., De Souza Chaves, I., et al. (2020) ‘Downregulation of a Mitochondrial NAD+Transporter (NDT2) Alters Seed Production and Germination in Arabidopsis’, Plant and Cell Physiology, 61(5), pp. 897–908. Available at: https://doi.org/10.1093/pcp/pcaa017.

Flütsch, S., Nigro, A., et al. (2020) ‘ Glucose uptake to guard cells via STP transporters provides carbon sources for stomatal opening and plant growth ’, EMBO reports, 21(8). Available at: https://doi.org/10.15252/embr.201949719.

Flütsch, S., Wang, Y., et al. (2020) ‘Guard cell starch degradation yields glucose for rapid stomatal opening in Arabidopsis’, Plant Cell, 32(7), pp. 2325–2344. Available at: https://doi.org/10.1105/tpc.18.00802.

Franks, P.J. et al. (2015) ‘Increasing water-use efficiency directly through genetic manipulation of stomatal density’, New Phytologist, 207(1), pp. 188–195. Available at: https://doi.org/10.1111/nph.13347.

Franks, P.J. and Beerling, D.J. (2009) ‘Maximum leaf conductance driven by CO2 effects on stomatal size and density over geologic time’, Proceedings of the National Academy of Sciences of the United States of America, 106(25), pp. 10343–10347. Available at: https://doi.org/10.1073/pnas.0904209106.

Gago, J. et al. (2019) ‘Photosynthesis Optimized across Land Plant Phylogeny’, Trends in Plant Science, pp. 947–958. Available at: https://doi.org/10.1016/j.tplants.2019.07.002.

Gao, H. et al. (2012) ‘Characterization of H2O-forming NADH oxidase from Streptococcus pyogenes and its application in l-rare sugar production’, Bioorganic and Medicinal Chemistry Letters, 22(5), pp. 1931–1935. Available at: https://doi.org/10.1016/j.bmcl.2012.01.049.

Gao, H. et al. (2019) ‘NADH oxidase from Lactobacillus reuteri: A versatile enzyme for oxidized cofactor regeneration’, International Journal of Biological Macromolecules, 123, pp. 629–636. Available at: https://doi.org/10.1016/j.ijbiomac.2018.11.096.

Geueke, B., Riebel, B. and Hummel, W. (2003) ‘NADH oxidase from Lactobacillus brevis: A new catalyst for the regeneration of NAD’, Enzyme and Microbial Technology, 32(2), pp. 205–211. Available at: https://doi.org/10.1016/S0141-0229(02)00290-9.

de Gonzalo, G., A. Orden, A. and R. Bisogno, F. (2012) ‘New Trends in Organic Synthesis with Oxidative Enzymes’, Current Organic Chemistry, 16(21), pp. 2598–2612. Available at: https://doi.org/10.2174/138527212804004599.

Grützner, R. et al. (2021) ‘High-efficiency genome editing in plants mediated by a Cas9 gene containing multiple introns’, Plant Communications, 2(2). Available at: https://doi.org/10.1016/j.xplc.2020.100135.

Gudesblat, G.E. et al. (2012) ‘SPEECHLESS integrates brassinosteroid and stomata signalling pathways’, Nature Cell Biology, 14(5), pp. 548–554. Available at: https://doi.org/10.1038/ncb2471.

Hara, K. et al. (2007) ‘The secretory peptide gene EPF1 enforces the stomatal one-cell-spacing rule’, Genes and Development, 21(14), pp. 1720–1725. Available at: https://doi.org/10.1101/gad.1550707.

Hara, K. et al. (2009) ‘Epidermal cell density is autoregulated via a secretory peptide, EPIDERMAL PATTERNING FACTOR 2 in arabidopsis leaves’, Plant and Cell Physiology, 50(6), pp. 1019–1031. Available at: https://doi.org/10.1093/pcp/pcp068.

Harrison, E.L. et al. (2020) ‘The influence of stomatal morphology and distribution on photosynthetic gas exchange’, Plant Journal, 101(4), pp. 768–779. Available at: https://doi.org/10.1111/tpj.14560.

Hashida, S.N. et al. (2007) ‘Arabidopsis thaliana nicotinate/nicotinamide mononucleotide adenyltransferase (AtNMNAT) is required for pollen tube growth’, Plant Journal, 49(4), pp. 694–703. Available at: https://doi.org/10.1111/j.1365-313X.2006.02989.x.

Hashida, S.N. et al. (2010) ‘Nicotinate/nicotinamide mononucleotide adenyltransferase-mediated regulation of NAD biosynthesis protects guard cells from reactive oxygen species in ABA-mediated stomatal movement in Arabidopsis’, Journal of Experimental Botany, 61(13), pp. 3813–3825. Available at: https://doi.org/10.1093/jxb/erq190.

Hashida, S.N. et al. (2013) ‘NAD+ accumulation during pollen maturation in arabidopsis regulating onset of germination’, Molecular Plant, 6(1), pp. 216–225. Available at: https://doi.org/10.1093/mp/sss071.

Hashida, S.N., Kawai-Yamada, M. and Uchimiya, H. (2013) ‘NAD+ accumulation as a metabolic off switch for orthodox pollen’, Plant Signaling and Behavior, 8(5). Available at: https://doi.org/10.4161/psb.23937.

Haworth, M., Elliott-Kingston, C. and McElwain, J.C. (2011) ‘Stomatal control as a driver of plant evolution’, Journal of Experimental Botany, 62(8), pp. 2419–2423. Available at: https://doi.org/10.1093/jxb/err086.

Hepworth, C. et al. (2015) 'Manipulating stomatal density enhances drought tolerance without deleterious effect on nutrient uptake', New Phytologist, 208(2), pp. 336–341. Available at: https://doi.org/10.1111/nph.13598.

Heux, S., Cachon, R. and Dequin, S. (2006) 'Cofactor engineering in Saccharomyces cerevisiae: Expression of a H2O-forming NADH oxidase and impact on redox metabolism', Metabolic Engineering, 8(4), pp. 303–314. Available at: https://doi.org/10.1016/j.ymben.2005.12.003.

Higuchi, M. et al. (1993) 'Identification of two distinct NADH oxidases corresponding to H2O2-forming oxidase and H2O-forming oxidase induced in Streptococcus mutans', Journal of General Microbiology, 139(10), pp. 2343–2351. Available at: https://doi.org/10.1099/00221287-139-10-2343.

Horaruang, W. et al. (2022) 'Engineering a K+ channel "sensory antenna" enhances stomatal kinetics, water use efficiency and photosynthesis', Nature Plants, 8(11), pp. 1262–1274. Available at: https://doi.org/10.1038/s41477-022-01255-2.

Horrer, D. et al. (2016) 'Blue light induces a distinct starch degradation pathway in guard cells for stomatal opening', Current Biology, 26(3), pp. 362–370. Available at: https://doi.org/10.1016/j.cub.2015.12.036.

Huang, S. et al. (2021) 'Optogenetic control of the guard cell membrane potential and stomatal movement by the light-gated anion channel GtACR1', Science Advances, 7(28). Available at: https://doi.org/10.1126/sciadv.abg4619.

Hughes, J. et al. (2017) 'Reducing stomatal density in barley improves drought tolerance without impacting on yield', Plant Physiology, 174(2), pp. 776–787. Available at: https://doi.org/10.1104/pp.16.01844.

Hummel, W. and Riebel, B. (2003) 'Isolation and biochemical characterization of a new NADH oxidase from Lactobacillus brevis', Biotechnology Letters, 25(1), pp. 51–54. Available at: https://doi.org/10.1023/A:1021730131633.

Igamberdiev, A.U. and Gardeström, P. (2003) 'Regulation of NAD- and NADP-dependent isocitrate dehydrogenases by reduction levels of pyridine nucleotides in mitochondria and cytosol of pea leaves', Biochimica et Biophysica Acta - Bioenergetics, 1606(1–3), pp. 117–125. Available at: https://doi.org/10.1016/S0005-2728(03)00106-3.

Jalakas, P. et al. (2018) 'ABA-mediated regulation of stomatal density is OST1-independent', Plant Direct, 2(9). Available at: https://doi.org/10.1002/pld3.82.

Jia, B. et al. (2008) 'Hexameric ring structure of a thermophilic archaeon NADH oxidase that produces predominantly H2O', FEBS Journal, 275(21), pp. 5355–5366. Available at: https://doi.org/10.1111/j.1742-4658.2008.06665.x.

Jia, B. et al. (2010) 'An archaeal NADH oxidase causes damage to both proteins and nucleic acids under oxidative stress.', Molecules and cells, 29(4), pp. 363–371. Available at: https://doi.org/10.1007/s10059-010-0045-8.

Katoh, A. et al. (2006) 'Early steps in the biosynthesis of NAD in arabidopsis start with aspartate and occur in the plastid', Plant Physiology, 141(3), pp. 851–857. Available at: https://doi.org/10.1104/pp.106.081091.

Kawasaki, S. et al. (2004) 'Purification and characterization of an H2O-forming NADH oxidase from Clostridium aminovalericum: Existence of an oxygen-detoxifying enzyme in an obligate anaerobic bacteria', Archives of Microbiology, 181(4), pp. 324–330. Available at: https://doi.org/10.1007/s00203-004-0659-3.

Kelly, G. et al. (2012) 'The pitfalls of transgenic selection and new roles of AtHXK1: A high level of AtHXK1 expression uncouples hexokinase1 dependent sugar signaling from exogenous sugar', Plant Physiology, 159(1), pp. 47–51. Available at: https://doi.org/10.1104/pp.112.196105.

Kelly, G. et al. (2019) 'Guard-Cell Hexokinase Increases Water-Use Efficiency Under Normal and Drought Conditions', Frontiers in Plant Science, 10. Available at: https://doi.org/10.3389/fpls.2019.01499.

Kondo, T. et al. (2010) 'Stomatal density is controlled by a mesophyll-derived signaling molecule', Plant and Cell Physiology, 51(1), pp. 1–8. Available at: https://doi.org/10.1093/pcp/pcp180.

Kroutil, W. et al. (2004) 'Biocatalytic Oxidation of Primary and Secondary Alcohols', ChemInform, 35(27). Available at: https://doi.org/10.1002/chin.200427225.

Kusumi, K. et al. (2012) 'Increased leaf photosynthesis caused by elevated stomatal conductance in a rice mutant deficient in SLAC1, a guard cell anion channel protein', Journal of Experimental Botany, 63(15), pp. 5635–5644. Available at: https://doi.org/10.1093/jxb/ers216.

Lake, J.A. and Woodward, F.I. (2008) 'Response of stomatal numbers to CO2 and humidity: Control by transpiration rate and abscisic acid', New Phytologist, 179(2), pp. 397–404. Available at: https://doi.org/10.1111/j.1469-8137.2008.02485.x.

Laporte, M.M., Shen, B. and Tarczynski, M.C. (2002) 'Engineering for drought avoidance: Expression of maize NADP-malic enzyme in tobacco results in altered stomatal function', Journal of Experimental Botany, 53(369), pp. 699–705. Available at: https://doi.org/10.1093/jexbot/53.369.699.

Lau, O.S. et al. (2018) 'Direct Control of SPEECHLESS by PIF4 in the High-Temperature Response of Stomatal Development', Current Biology, 28(8), pp. 1273-1280.e3. Available at: https://doi.org/10.1016/j.cub.2018.02.054.

Lawson, T. and Blatt, M.R. (2014) 'Stomatal size, speed, and responsiveness impact on photosynthesis and water use efficiency', Plant Physiology, 164(4), pp. 1556–1570. Available at: https://doi.org/10.1104/pp.114.237107.

Leakey, A.D.B. et al. (2019) 'Water Use Efficiency as a Constraint and Target for Improving the Resilience and Productivity of C 3 and C 4 Crops', Annual Review of Plant Biology, 70, pp. 781–808. Available at: https://doi.org/10.1146/annurev-arplant-042817-040305.

Li, F.L. et al. (2018) 'Cloning, expression, characterization and homology modeling of a novel water-forming NADH oxidase from Streptococcus mutans ATCC 25175', International Journal of Biological Macromolecules, 113, pp. 1073–1079. Available at: https://doi.org/10.1016/j.ijbiomac.2018.03.016.

Li, F.L. et al. (2019) 'Switching the substrate specificity from NADH to NADPH by a single mutation of NADH oxidase from Lactobacillus rhamnosus', International Journal of Biological Macromolecules, 135, pp. 328–336. Available at: https://doi.org/10.1016/j.ijbiomac.2019.05.146.

Li, L. and Wang, C.C. (2006) 'A likely molecular basis of the susceptibility of Giardia lamblia towards oxygen', Molecular Microbiology, 59(1), pp. 202–211. Available at: https://doi.org/10.1111/j.1365-2958.2005.04896.x.

López-Calcagno, P.E. et al. (2020) 'Stimulating photosynthetic processes increases productivity and water-use efficiency in the field', Nature Plants, 6(8), pp. 1054–1063. Available at: https://doi.org/10.1038/s41477-020-0740-1.

Lopez de Felipe, F. and Hugenholtz, J. (2001) 'Purification and characterisation of the water forming NADH-oxidase from Lactococcus lactis', International Dairy Journal, 11(1–2), pp. 37–44. Available at: https://doi.org/10.1016/S0958-6946(01)00031-0.

Lountos, G.T. et al. (2006) ‘The crystal structure of NAD(P)H oxidase from Lactobacillus sanfranciscensis: Insights into the conversion of O2 into two water molecules by the flavoenzyme’, Biochemistry, 45(32), pp. 9648–9659. Available at: https://doi.org/10.1021/bi060692p.

Lugassi, N. et al. (2015) ‘Expression of Arabidopsis hexokinase in citrus guard cells controls stomatal aperture and reduces transpiration’, Frontiers in plant science, 6.

Lugassi, N. et al. (2019) ‘Expression of Arabidopsis hexokinase in tobacco guard cells increases water-use efficiency and confers tolerance to drought and salt stress’, Plants, 8(12). Available at: https://doi.org/10.3390/plants8120613.

Lundgren, M.R. et al. (2019) ‘Mesophyll porosity is modulated by the presence of functional stomata’, Nature Communications, 10(1). Available at: https://doi.org/10.1038/s41467-019-10826-5.

Ma, S.H. et al. (2021) ‘The ten amino acids of the oxygen-evolving enhancer of tobacco is sufficient as the peptide residues for protein transport to the chloroplast thylakoid’, Plant Molecular Biology, 105(4–5), pp. 513–523. Available at: https://doi.org/10.1007/s11103-020-01106-8.

Marillonnet, S. and Engler, C. (2014) ‘Golden Gate Cloning’, Methods in Molecular Biology, 1116, pp. 119–131. Available at: http://link.springer.com/10.1007/978-1-62703-764-8.

Masle, J., Gilmore, S.R. and Farquhar, G.D. (2005) ‘The ERECTA gene regulates plant transpiration efficiency in Arabidopsis’, Nature, 436(7052), pp. 866–870. Available at: https://doi.org/10.1038/nature03835.

Matsumoto, J. et al. (1996) ‘Molecular cloning and sequence analysis of the gene encoding the H2O-forming NADH Oxidase from streptococcus mutans’, Bioscience, Biotechnology and Biochemistry, 60(1), pp. 39–43. Available at: https://doi.org/10.1271/bbb.60.39.

McAusland, L. et al. (2016) ‘Effects of kinetics of light-induced stomatal responses on photosynthesis and water-use efficiency’, The New phytologist, 211(4), pp. 1209–1220. Available at: https://doi.org/10.1111/nph.14000.

McElwain, J.C. and Chaloner, W.G. (1995) ‘Stomatal density and index of fossil plants track atmospheric carbon dioxide in the palaeozoic’, Annals of Botany, 76(4), pp. 389–395. Available at: https://doi.org/10.1006/anbo.1995.1112.

Mcelwain, J.C., Yiotis, C. and Lawson, T. (2016) ‘Using modern plant trait relationships between observed and theoretical maximum stomatal conductance and vein density to examine patterns of plant macroevolution’, New Phytologist, 209(1), pp. 94–103. Available at: https://doi.org/10.1111/nph.13579.

Medeiros, D.B. et al. (2016) ‘Enhanced photosynthesis and growth in atquac1 knockout mutants are due to altered organic acid accumulation and an increase in both stomatal and mesophyll conductance’, Plant Physiology, 170(1), pp. 86–101. Available at: https://doi.org/10.1104/pp.15.01053.

Medeiros, D.B. et al. (2017) ‘Impaired malate and fumarate accumulation due to the mutation of the tonoplast dicarboxylate transporter has little effects on stomatal behavior’, Plant Physiology, 175(3), pp. 1068–1081. Available at: https://doi.org/10.1104/pp.17.00971.

Mitra, A., Choi, H.K. and An, G. (1989) ‘Structural and functional analyses of Arabidopsis thaliana chlorophyll a/b-binding protein (cab) promoters’, Plant Molecular Biology, 12(2), pp. 169–179. Available at: https://doi.org/10.1007/BF00020502.

Mohammed, U. et al. (2019) 'Rice plants overexpressing OsEPF1 show reduced stomatal density and increased root cortical aerenchyma formation', Scientific Reports, 9(1). Available at: https://doi.org/10.1038/s41598-019-41922-7.

Müller, G.L. et al. (2018) 'Improved water use efficiency and shorter life cycle of Nicotiana tabacum due to modification of guard and vascular companion cells', Scientific Reports, 8(1). Available at: https://doi.org/10.1038/s41598-018-22431-5.

Murray, M.G. and Thompson, W.F. (1980) 'Rapid isolation of high molecular weight plant DNA', Nucleic Acids Research, 8(19), pp. 4321–4326. Available at: https://doi.org/10.1093/nar/8.19.4321.

Nadeau, J.A. and Sack, F.D. (2002) 'Control of stomatal distribution on the Arabidopsis leaf surface', Science, 296(5573), pp. 1697–1700. Available at: https://doi.org/10.1126/science.1069596.

Nisar, M.A. et al. (2013) 'TK1299, a highly thermostable NAD(P)H oxidase from Thermococcus kodakaraensis exhibiting higher enzymatic activity with NADPH', Journal of Bioscience and Bioengineering, 116(1), pp. 39–44. Available at: https://doi.org/10.1016/j.jbiosc.2013.01.020.

Nowak, C. et al. (2015) 'A water-forming NADH oxidase from Lactobacillus pentosus suitable for the regeneration of synthetic biomimetic cofactors', Frontiers in Microbiology, 6(SEP). Available at: https://doi.org/10.3389/fmicb.2015.00957.

Papanatsiou, M. et al. (2019) 'Optogenetic manipulation of stomatal kinetics improves carbon assimilation,water use, and growth', Science, 363(6434), pp. 1456–1459. Available at: https://doi.org/10.1126/science.aaw0046.

Papanatsiou, M., Amtmann, A. and Blatt, M.R. (2016) 'Stomatal spacing safeguards stomatal dynamics by facilitating guard cell ion transport independent of the epidermal solute reservoir', Plant Physiology, 172(1), pp. 254–263. Available at: https://doi.org/10.1104/pp.16.00850.

Park, J.T. et al. (2011) 'NAD(P)H oxidase v from Lactobacillus plantarum (NoxV) displays enhanced operational stability even in absence of reducing agents', Journal of Molecular Catalysis B: Enzymatic, 71(3–4), pp. 159–165. Available at: https://doi.org/10.1016/j.molcatb.2011.04.013.

Penfield, S. et al. (2012) 'Expression and manipulation of PHOSPHOENOLPYRUVATE CARBOXYKINASE 1 identifies a role for malate metabolism in stomatal closure', Plant Journal, 69(4), pp. 679–688. Available at: https://doi.org/10.1111/j.1365-313X.2011.04822.x.

Pérez-Bueno, M.L. et al. (2022) 'An extremely low stomatal density mutant overcomes cooling limitations at supra-optimal temperature by adjusting stomatal size and leaf thickness', Frontiers in Plant Science, 13. Available at: https://doi.org/10.3389/fpls.2022.919299.

Petschacher, B. et al. (2014) 'Cofactor specificity engineering of streptococcus mutans NADH oxidase 2 for nad(p)+ regeneration in biocatalytic oxidations', Computational and Structural Biotechnology Journal, 9(14), p. e201402005. Available at: https://doi.org/10.5936/csbj.201402005.

Pillitteri, L.J. et al. (2007) 'Termination of asymmetric cell division and differentiation of stomata', Nature, 445(7127), pp. 501–505. Available at: https://doi.org/10.1038/nature05467.

Prasch, C.M. et al. (2015) 'β-amylase1 mutant Arabidopsis plants show improved drought tolerance due to reduced starch breakdown in guard cells', Journal of Experimental Botany, 66(19), pp. 6059–6067. Available at: https://doi.org/10.1093/jxb/erv323.

Quarrie, S.A. and Jones, H.G. (1977) 'Effects of abscisic acid and water stress on development and morphology of wheat', Journal of Experimental Botany, 28(1), pp. 192–203. Available at: https://doi.org/10.1093/jxb/28.1.192.

Rashid, N. et al. (2011) ‘Gene cloning and characterization of NADH oxidase from thermococcus kodakarensis’, African Journal of Biotechnology, 10(78), pp. 17916–17924. Available at: https://doi.org/10.5897/AJB11.989.

Raven, J.A. (2002) ‘Selection pressures on stomatal evolution’, New Phytologist, 153(3), pp. 371–386. Available at: https://doi.org/10.1046/j.0028-646X.2001.00334.x.

Reyer, A. et al. (2020) ‘Channelrhodopsin-mediated optogenetics highlights a central role of depolarization-dependent plant proton pumps’, Proceedings of the National Academy of Sciences of the United States of America, 117(34), pp. 20920–20925. Available at: https://doi.org/10.1073/PNAS.2005626117.

Riebel, B.R. et al. (2002) Cofactor Regeneration of NAD+ from NADH: Novel Water-Forming NADH Oxidases, Advanced Synthesis and Catalysis. Available at: https://doi.org/10.1002/1615-4169(200212)344:10<1156::aid-adsc1156>3.0.co;2-%23.

Sakamoto, M., Uchimura, T. and Komagata, K. (1996) ‘Comparison of H2O-Forming NADH Oxidase from Leuconostoc mesenteroides subsp. mesenteroides NRIC 1541T and H2O2-Forming NADH Oxidase from Sporolactobacillus inulinus NRIC 1133T’, Journal of Fermentation and Bioengineering, 82(6), pp. 531–537. Available at: https://doi.org/10.1016/S0922-338X(97)81247-0.

Sakoda, K. et al. (2020) ‘Higher Stomatal Density Improves Photosynthetic Induction and Biomass Production in Arabidopsis Under Fluctuating Light’, Frontiers in Plant Science, 11. Available at: https://doi.org/10.3389/fpls.2020.589603.

Savaldi-Goldstein, S., Peto, C. and Chory, J. (2007) ‘The epidermis both drives and restricts plant shoot growth’, Nature, 446(7132), pp. 199–202. Available at: https://doi.org/10.1038/nature05618.

Schlüter, U. et al. (2003) ‘Photosynthetic performance of an Arabidopsis mutant with elevated stomatal density (sdd1-1) under different light regimes’, Journal of Experimental Botany, 54(383), pp. 867–874. Available at: https://doi.org/10.1093/jxb/erg087.

Schmidt, H. L et al. (1986) ‘Isolation and properties of an H2O-forming NADH oxidase from Streptococcus faecalis’, European Journal of Biochemistry, 156(1), pp. 149–155. Available at: https://doi.org/10.1111/j.1432-1033.1986.tb09560.x.

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) ‘NIH Image to ImageJ: 25 years of image analysis’, Nature Methods, 9(7), pp. 671–675. Available at: https://doi.org/10.1038/nmeth.2089.

Schuler, M.L. et al. (2018) ‘SHORTROOT-mediated increase in stomatal density has no impact on photosynthetic efficiency’, Plant Physiology, 176(1), pp. 757–772. Available at: https://doi.org/10.1104/pp.17.01005.

Serna, L. and Fenoll, C. (2000) ‘Stomatal development in Arabidopsis: How to make a functional pattern’, Trends in Plant Science, 5(11), pp. 458–460. Available at: https://doi.org/10.1016/S1360-1385(00)01782-9.

Shpak, E.D. et al. (2005) ‘Stomatal patterning and differentiation by synergistic interactions of receptor kinases’, Science, 309(5732), pp. 290–293. Available at: https://doi.org/10.1126/science.1109710.

Smith, E.N. et al. (2021) ‘Shining a light on NAD- and NADP-based metabolism in plants’, Trends in Plant Science [Preprint]. Available at: https://doi.org/10.1016/j.tplants.2021.06.010.

de Souza Chaves, I. et al. (2019) ‘The mitochondrial NAD+ transporter (NDT1) plays important roles in cellular NAD+ homeostasis in Arabidopsis thaliana’, Plant Journal, 100(3), pp. 487–504. Available at: https://doi.org/10.1111/tpj.14452.

Stanton, T.B. and Jensen, N.S. (1993) 'Purification and characterization of NADH oxidase from Serpulina (Treponema) hyodysenteriae', Journal of Bacteriology, 175(10), pp. 2980–2987. Available at: https://doi.org/10.1128/jb.175.10.2980-2987.1993.

Sun, L. et al. (2017) 'NADK2 positively modulates abscisic acid-induced stomatal closure by affecting accumulation of H2O2, Ca2+ and nitric oxide in Arabidopsis guard cells', Plant Science, 262, pp. 81–90. Available at: https://doi.org/10.1016/j.plantsci.2017.06.003.

Tanaka, Yoko et al. (2013) 'ABA inhibits entry into stomatal-lineage development in Arabidopsis leaves', Plant Journal, 74(3), pp. 448–457. Available at: https://doi.org/10.1111/tpj.12136.

Tanaka, Yu et al. (2013) 'Enhancement of leaf photosynthetic capacity through increased stomatal density in Arabidopsis', New Phytologist, 198(3), pp. 757–764. Available at: https://doi.org/10.1111/nph.12186.

Titov, D. V. et al. (2016) 'Complementation of mitochondrial electron transport chain by manipulation of the NAD+/NADH ratio', Science, 352(6282), pp. 231–235. Available at: https://doi.org/10.1126/science.aad4017.

Toh, S. et al. (2021) 'Overexpression of Plasma Membrane H+-ATPase in Guard Cells Enhances Light-Induced Stomatal Opening, Photosynthesis, and Plant Growth in Hybrid Aspen', Frontiers in Plant Science, 12. Available at: https://doi.org/10.3389/fpls.2021.766037.

Turner, N.J. (2011) 'Enantioselective Oxidation of C-O and C-N bonds using oxidases', Chemical Reviews, 111(7), pp. 4073–4087. Available at: https://doi.org/10.1021/cr200111v.

Vemuri, G.N. et al. (2007) 'Increasing NADH oxidation reduces overflow metabolism in Saccharomyces cerevisiae', Proceedings of the National Academy of Sciences of the United States of America, 104(7), pp. 2402–2407. Available at: https://doi.org/10.1073/pnas.0607469104.

Vialet-Chabrand, S., Dreyer, E. and Brendel, O. (2013) 'Performance of a new dynamic model for predicting diurnal time courses of stomatal conductance at the leaf level', Plant, Cell and Environment, 36(8), pp. 1529–1546. Available at: https://doi.org/10.1111/pce.12086.

Vile, D. et al. (2012) 'Arabidopsis growth under prolonged high temperature and water deficit: Independent or interactive effects?', Plant, Cell and Environment, 35(4), pp. 702–718. Available at: https://doi.org/10.1111/j.1365-3040.2011.02445.x.

Wang, C. et al. (2016) 'PdEPF1 regulates water-use efficiency and drought tolerance by modulating stomatal density in poplar', Plant Biotechnology Journal, 14(3), pp. 849–860. Available at: https://doi.org/10.1111/pbi.12434.

Wang, H. et al. (1998) 'ICK1, a cyclin-dependent protein kinase inhibitor from Arabidopsis thaliana interacts with both Cdc2a and CycD3, and its expression is induced by abscisic acid', Plant Journal, 15(4), pp. 501–510. Available at: https://doi.org/10.1046/j.1365-313X.1998.00231.x.

Wang, H. et al. (2007) 'Stomatal development and patterning are regulated by environmentally responsive mitogen-activated protein kinases in Arabidopsis', Plant Cell, 19(1), pp. 63–73. Available at: https://doi.org/10.1105/tpc.106.048298.

Wang, H. et al. (2019) 'A subsidiary cell-localized glucose transporter promotes stomatal conductance and photosynthesis', Plant Cell, 31(6), pp. 1328–1343. Available at: https://doi.org/10.1105/tpc.18.00736.

Wang, L., Chong, H. and Jiang, R. (2012) 'Comparison of alkyl hydroperoxide reductase and two water-forming NADH oxidases from Bacillus cereus ATCC 14579', Applied Microbiology and Biotechnology, 96(5), pp. 1265–1273. Available at: https://doi.org/10.1007/s00253-012-3919-1.

Wang, Y. et al. (2014) 'Overexpression of plasma membrane H+-ATPase in guard cells promotes light-induced stomatal opening and enhances plant growth', Proceedings of the National Academy of Sciences of the United States of America, 111(1), pp. 533–538. Available at: https://doi.org/10.1073/pnas.1305438111.

Ward, D.E. et al. (2001) 'The NADH oxidase from Pyrococcus furiosus: Implications for the protection of anaerobic hyperthermophiles against oxidative stress', European Journal of Biochemistry, 268(22), pp. 5816–5823. Available at: https://doi.org/10.1046/j.0014-2956.2001.02526.x.

Wieloch, T. and Sharkey, T.D. (2022) 'Compartment-specific energy requirements of photosynthetic carbon metabolism in Camelina sativa leaves', Planta, 255(5). Available at: https://doi.org/10.1007/s00425-022-03884-5.

Woodward, F.I. (1987) 'Stomatal numbers are sensitive to increases in CO2 from pre-industrial levels', Nature, 327, pp. 617–618.

Woodward, F.I. and Bazzaz, F.A. (1988) 'The responses of stomatal density to CO2 partial pressure', Journal of Experimental Botany, 39(12), pp. 1771–1781. Available at: https://doi.org/10.1093/jxb/39.12.1771.

Woodward, F.I., Lake, J.A. and Quick, W.P. (2002) 'Stomatal development and CO2: Ecological consequences', New Phytologist, 153(3), pp. 477–484. Available at: https://doi.org/10.1046/j.0028-646X.2001.00338.x.

Wu, F.H. et al. (2009) 'Tape-arabidopsis sandwich - A simpler arabidopsis protoplast isolation method', Plant Methods, 5(1). Available at: https://doi.org/10.1186/1746-4811-5-16.

Wu, S. and Zhao, B. (2017) 'Using clear nail polish to make Arabidopsis epidermal impressions for measuring the change of stomatal aperture size in immune response', in Methods in Molecular Biology, pp. 243–248. Available at: https://doi.org/10.1007/978-1-4939-6859-6_20.

Wu, X. et al. (2012) 'Application of a novel thermostable NAD(P)H oxidase from hyperthermophilic archaeon for the regeneration of both NAD + and NADP +', Biotechnology and Bioengineering, 109(1), pp. 53–62. Available at: https://doi.org/10.1002/bit.23294.

Yan, M. et al. (2016) 'Characteristics of a water-forming NADH oxidase from Methanobrevibacter smithii, an archaeon in the human gut', Bioscience Reports, 36(6). Available at: https://doi.org/10.1042/BSR20160357.

Yang, M. and Sack, F.D. (1995) 'The too many mouths and four lips mutations affect stomatal production in arabidopsis', Plant Cell, 7(12), pp. 2227–2239. Available at: https://doi.org/10.2307/3870164.

Yang, X. et al. (2022) 'Abscisic acid regulates stomatal production by imprinting a SnRK2 kinase–mediated phosphocode on the master regulator SPEECHLESS', Science Advances, 8(40). Available at: https://doi.org/10.1126/sciadv.add2063.

Yin, X. et al. (2017) 'CRISPR-Cas9 and CRISPR-Cpf1 mediated targeting of a stomatal developmental gene EPFL9 in rice', Plant Cell Reports, 36(5), pp. 745–757. Available at: https://doi.org/10.1007/s00299-017-2118-z.

Yoo, C.Y. et al. (2010) 'The Arabidopsis GTL1 transcription factor regulates water use efficiency and drought tolerance by modulating stomatal density via transrepression of SDD1', Plant Cell, 22(12), pp. 4128–4141. Available at: https://doi.org/10.1105/tpc.110.078691.

Yoo, S.D., Cho, Y.H. and Sheen, J. (2007) 'Arabidopsis mesophyll protoplasts: A versatile cell system for transient gene expression analysis', Nature Protocols, 2(7), pp. 1565–1572. Available at: https://doi.org/10.1038/nprot.2007.199.

Yu, Q. et al. (2020) 'RSD1 Is Essential for Stomatal Patterning and Files in Rice', Frontiers in Plant Science, 11. Available at: https://doi.org/10.3389/fpls.2020.600021.

Zhang, Y.W. et al. (2012) 'Cloning and characterization of a thermostable H2O-forming NADH oxidase from Lactobacillus rhamnosus', Enzyme and Microbial Technology, 50(4–5), pp. 255–262. Available at: https://doi.org/10.1016/j.enzmictec.2012.01.009.

Zhao, J. et al. (2024) 'Genome-wide analysis of the TIFY family in Lycium and the negative regulation of stomatal development by LrJAZ2 gene', Plant Physiology and Biochemistry, 206. Available at: https://doi.org/10.1016/j.plaphy.2023.108285.

Zhou, Y. et al. (2021) 'Optogenetic control of plant growth by a microbial rhodopsin', Nature Plants, 7(2), pp. 144–151. Available at: https://doi.org/10.1038/s41477-021-00853-w.

Chapter 5. General Discussion

Chapter 5, Contributors:
Bouvier, J.W. and Kelly, S.

Chapter 5, Author contributions:
J.W.B. conceptualisation, reading, and writing. S.K. edits, comments, and discussion.

5.1 Thesis key findings

In this body of work, I have integrated insights from evolution and genetic engineering to advance our understanding of CO_2 assimilation in plants and explore the potential for its future optimisation. This research has yielded a number of key findings that each address one of the core aims set out at the start of this thesis. These key findings are summarised below. The broader implications of these findings are discussed in dedicated discussion Chapters 5.2 – 5.5.

The first research stream in this thesis focuses on the enzyme rubisco as the primary biochemical limitation on plant CO_2 assimilation and productivity. In Chapter 2, I identify a strong phylogenetic signal in rubisco kinetic data and show that kinetic trait trade-offs are weaker than previously thought when correcting for this signal. I also make the novel discovery that phylogenetic constraints have provided a more severe limitation on rubisco adaptation compared to the combined action of all kinetic trait trade-offs. In Chapter 3, I extend the above analyses to perform the first contextualized investigation of rubisco molecular and kinetic evolution. This work demonstrates that the rubisco catalytic large subunit is among the 1% of slowest-evolving genes and enzymes on Earth. Importantly, this analysis also finds that rubisco is continually evolving for enhanced catalysis and leaf-level photosynthesis in plants. Combined, these chapters discover that rubisco kinetic traits have been evolving largely independently of one another in an adaptive landscape that is predominantly limited by phylogenetic constraint. Crucially, they also find that within this constrained landscape, rubisco is not locked in stasis. Instead, rubisco has evolved, and is continuing to evolve, for improved CO_2 assimilation capacity. Together, these key findings address Aims 1 – 3 as set out at the start of

this thesis. In Chapter 5.3 below, I offer insights regarding Aim 4 based upon further interpretation of these findings.

The second research stream in this thesis focuses on stomata as the primary physical limitation on plant CO_2 assimilation and productivity. In Chapter 4, I demonstrate the design, build, and test of a novel metabolic engineering strategy to modulate stomatal patterning. Specifically, I show that stomatal densities can be reduced in the model plant *Arabidopsis* thaliana through targeted manipulation of $NAD(P)^+$/NAD(P)H ratios using a genetically encoded tool. I also demonstrate that engineered plants exhibit an enhanced photosynthetic water-use efficiency without any impairment of net CO_2 assimilation rate or growth under optimal conditions. Finally, I make the important discovery that these improvements in photosynthesis-water relations translate to increased plant growth and biomass accumulation under water-deficit conditions such as those experienced during drought conditions in the field. Together, this chapter thus provides a proof-of-concept strategy which could help enhance the productivity and climate-resilience of agricultural crops. It also unlocks novel avenues to engineer plant stomata in future. Together, these key findings address Aim 5 as set out at the start of this thesis.

5.2 Phylogeny provides a powerful framework to study enzyme function

Evolution by natural selection has shaped the biological diversity on Earth. As a consequence of a singular origin of life followed by an unbroken chain of descent with modification (Doolittle, 2000; Nutman *et al.*, 2016), closely related individuals share greater resemblance on average than more distantly related individuals (Blomberg and Garland, 2002; Blomberg, Garland and Ives, 2003). This hierarchical structure exists at all levels of biological organisation, from the organism through every sequential higher order taxonomic rank up to the three domains of archaeal, bacterial, and eukaryotic life.

The above dogma provides a framework to systematically classify life. As such, classification was historically achieved by evaluating the relative similarities among the morphological, anatomical,

behavioural, and life-history traits of different organisms (Simpson, 1959). However, whilst phenotypic trait variation generally follows the structure of the tree of life and provides a good proxy for ancestry (Figure 5.1; Scenario 1), this assumption is not robust in all traits in the real world (Freckleton, Harvey and Pagel, 2002; Blomberg, Garland and Ives, 2003; Losos, 2008; Kamilar and Cooper, 2013). Instead, several phenomena such as divergent selection (Figure 5.1; Scenario 2), rapid adaptive bursts (Figure 5.1; Scenario 3), and convergent evolution (Figure 5.1; Scenario 4) can degrade the inherent phylogenetic structure of biological trait data (Hansen and Martins, 1996; Kamilar and Cooper, 2013; Kelly, Grenyer and Scotland, 2014). Specifically, this is because these evolutionary forces can either cause resemblance among closely related individuals to appear independent of evolutionary distance (Figure 5.1; Scenario 2 and 3), or they can instead cause distantly related species to appear more similar to one another than closely related species (Figure 5.1; Scenario 4). Although genotype determines phenotype, the evolutionary landscape is thus complex, and these complexities can obscure the use of phenotypic traits for inferring evolutionary history.

In line with the limitations of trait-based systematic approaches, molecular phylogenetics have become widely adopted for reconstructing evolutionary relationships during the current computational and genetic era of biology. A major reason for this is because genetic material offers several major advantages for systematics compared to phenotypic traits, including its considerably higher information content, it being encoded by non-ambiguous and non-overlapping states (i.e., unique nucleotide or amino acid residues), and it being inherited more faithfully across generations (Brown, 2002). Moreover, molecular phylogenetics can also be used to interrogate evolutionary relationships among all taxa, including groups such as microscopic prokaryotes which have evaded previous interrogation on account of lacking obvious phenotypic distinction (Tessler ***et al.***, 2022). Accordingly, molecular phylogenetics thus offers a more powerful alternative to previous trait-based approaches and has risen to prominence as the gold-standard method for systematics.

Alongside its considerable advantages, the widespread adoption of molecular phylogenetics has also been facilitated by its standardized methodology (Yang and Rannala, 2012; Ajawatanawong, 2017). In brief, sequence data is first obtained for orthologous genes (i.e., genes derived by descent

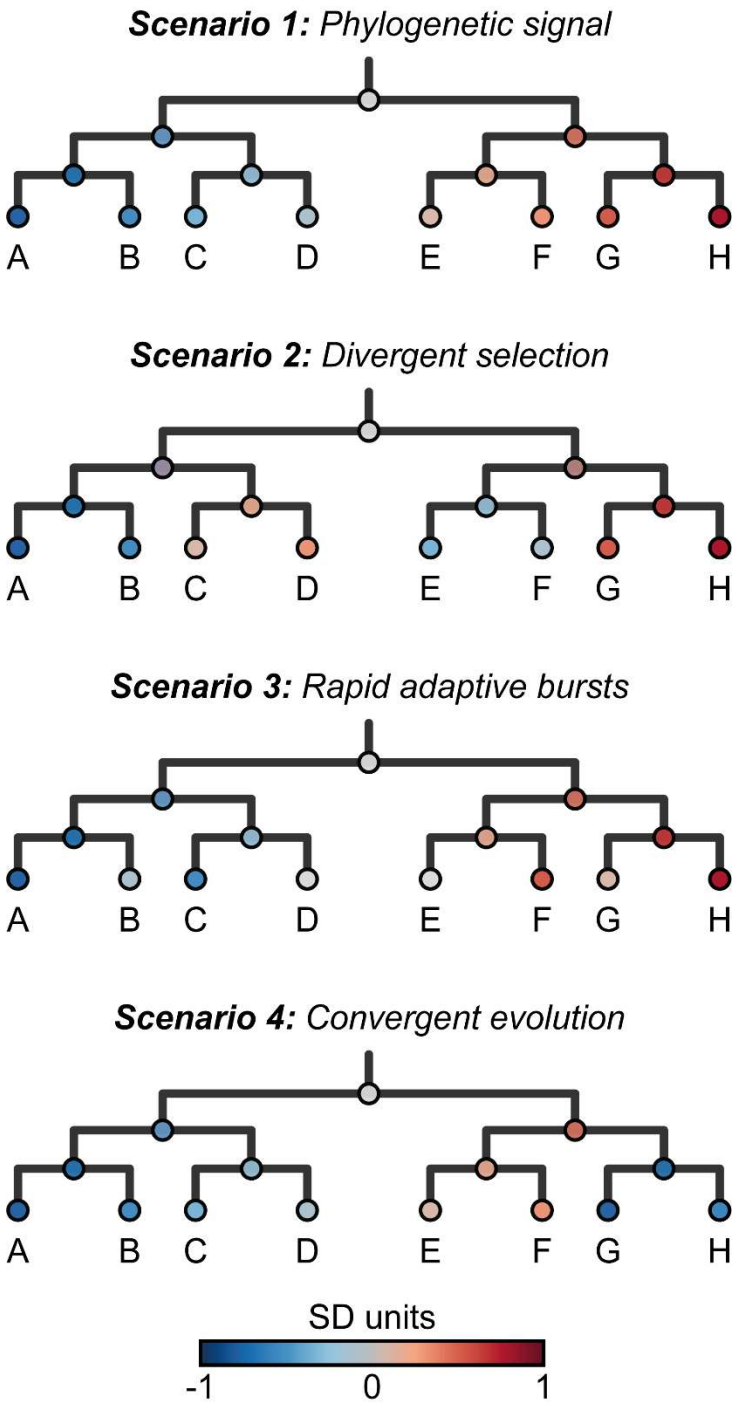


Figure 5.1. Cartoon to illustrate the phylogenetic structure in biological trait data under different evolutionary scenarios. Under all scenarios 1 – 4, the evolution of trait data is depicted by a colour gradient relative to an underlying molecular phylogeny which most correctly depicts the true species relationships. These are illustrated using eight extant species (identified by letters A—H). Scenario 1 illustrates the evolution of trait data as might be expected by inheritance and descent with modification. This depicts the null expectation of the presence of phylogenetic signal, in which more closely related species exhibit more similar trait values compared to more distantly related species. Scenario 2 – 4 illustrates the evolution of trait data as might be expected under divergent selection (Scenario 2), rapid adaptive bursts (Scenario 3), and convergent evolution (Scenario 4), respectively. All of these alternate scenarios depict a departure from phylogenetic signal in which either the

resemblance among closely related individuals appears independent of evolutionary distance (Scenario 2 and 3), or instead the resemblance among distantly related species appears more similar to one another than closely related species (Scenario 4). **[*End of Figure Legend*].**

---

from the same ancestral gene) across organisms. Next, these sequences are aligned by inserting gaps such that the most similar characters (nucleotides or amino acids) across species are contained in the same alignment position. Finally, mathematical algorithms are used to compute a phylogenetic tree which has a topology that proposes (either through model-based or distance-based approaches) the ancestry of all individual sequences with branch lengths that are proportionate to the amount of sequence change that has occurred between each parent-daughter node. Through this framework, molecular phylogenetics thus offers a robust and tractable strategy to infer hypotheses for the evolutionary relationships of sequences and species at a broad range of taxonomic and evolutionary scales.

As molecular phylogenetics enables inference of relationships for almost any biological process, the question arises: what extent do phenotypic traits conform to the tree-like pattern of evolution that would *a priori* be predicted from their underlying history? Answering this question is interesting from a fundamental biology perspective because it sheds light on the dynamics shaping trait evolution in the real world. Moreover, this is also of key relevance in fundamental comparative analyses to help avoid the “serious statistical problem” (Felsenstein, 1985) by which phylogenetically non-independent observations may drive spurious or incorrect associations. Of final importance, answering this question also has significance in diverse applied biology contexts including bioprospecting, genome mining, and emerging fields such as precision medicine which are all ultimately predicated upon the underlying assumption that phenotype tracks genotype. Thus, unravelling the concordance between phenotypic traits and the underlying molecular relationships of the species which encode these traits has significant and far-reaching implications in the biological sciences.

To interrogate the statistical co-variation between trait values and sequence similarity along a phylogenetic tree and thereby explore novel questions prompted by the advent of molecular

phylogenetics, a suite of comparative tools termed “phylogenetic signal” analyses have been developed (Gittleman and Kot, 1990; Abouheif, 1999; Pagel, 1999; Blomberg, Garland and Ives, 2003). In brief, these analyses collectively assess the presence and extent of correlation between species genotypes and phenotypes through computing a phylogenetic signal metric. The presence of strong phylogenetic signal means that trait values follow the structure of the underlying molecular phylogeny and the evolutionary relationships of the species in question (e.g., Figure 5.1; Scenario 1). Inversely, weak or absent phylogenetic signal is interpreted to mean that traits vary across species either largely, or entirely, independent of their shared ancestry (e.g., Figure 5.1; Scenarios 2 – 4). Ultimately, phylogenetic signal thus measures whether closely related individuals on average resemble each other more than distantly related individuals, as would be the *a priori* expectation from Darwinian evolution. In this way, studying phylogenetic signal provides a powerful approach to interpret patterns of biological diversity and to probe genotype-phenotype interactions in nature.

Since their inception, phylogenetic signal detection methods have been routinely utilised to explore the structure and pattern of trait variation in disciplines of ecology (e.g., Wu *et al.*, 2021), animal behaviour (e.g., Balasubramaniam *et al.*, 2012), anthropology (e.g., Lukas, Towner and Borgerhoff Mulder, 2021), and conservation science (e.g., Fritz and Purvis, 2010). However, in contrast to these whole-organismal traits, comparatively little attention has been paid into investigating the presence of phylogenetic signal in traits at molecular resolutions. Notably, one particular application of phylogenetic signal that has been overlooked relates to the functional traits of enzymes.

Given that sequence variation determines function, the physiochemical properties across enzymes should in theory be determined by their shared inheritance. However, enzymes pose a particularly interesting case-study in this regard. This is because single point mutations have routinely been demonstrated to cause dramatic departures in the catalytic properties of enzymes (Cleton-Jansen *et al.*, 1991; Villar *et al.*, 1997; Johnson *et al.*, 2001). On the other hand, highly divergent enzymes in nature can also exhibit nearly interchangeable physiochemical properties despite having evolved in isolation over considerable periods of geological time (Espadaler *et al.*, 2008). Accordingly, deciphering the molecular basis of enzyme properties, as well as predicting the consequence of novel mutations on enzyme function, have proven extremely challenging tasks that have long

frustrated scientists (Minshull *et al.*, 2005; Damborsky and Brezovsky, 2014; Siegel *et al.*, 2015; Carlin *et al.*, 2016). This frustration has been particularly pronounced among scientists seeking to engineer enzymes, with the most famous example being that of the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) which has to date largely resisted improvement despite this representing a long-standing goal of plant science research (Prywes *et al.*, 2022; Gionfriddo, Zang and Hayer-Hartl, 2023). Thus, the dynamics of enzyme functional adaptation are complex, and accordingly, whether enzyme properties conform to the canonical phylogenetic nested structure which generally apply at other scales of biological organisation remains to be determined.

In an effort to address these missing gaps in our understanding of enzyme evolution, in Chapter 2 and Appendix A of this thesis I performed a phylogenetic signal analysis of enzyme functional traits (Bouvier *et al.*, 2021; Bouvier and Kelly, 2023). Specifically, I sought to evaluate the presence and magnitude of phylogenetic signal in rubisco kinetics (Bouvier *et al.*, 2021; Bouvier and Kelly, 2023). This enzyme represents a powerful model system as the basis of these analyses because it is one of the most widely sequenced and biochemically characterised proteins on the planet, owing to both its widespread utilisation in systematics (Gielly and Taberlet, 1994; APG, 1998, 2016) and its central importance as the principal carbon-fixing enzyme in photosynthesis (Farquhar, von Caemmerer and Berry, 1980), respectively. In this thesis, I discovered that there was a strong phylogenetic signal in all rubisco kinetic traits (Bouvier *et al.*, 2021; Bouvier and Kelly, 2023). Moreover, I found that the presence of this phylogenetic signal was pervasive at all scales across the tree of life, including among C_3 angiosperms, among all angiosperms, and among all photosynthetic organisms under investigation, respectively (Bouvier *et al.*, 2021; Bouvier and Kelly, 2023). Building upon important and earlier seminal studies of rubisco (Galmés *et al.*, 2014, 2015; Hermida-Carrera, Kapralov and Galmés, 2016; Orr *et al.*, 2016), this represents an important advancement as it embodies one of very few (and the largest) phylogenetically resolved interrogations of enzyme functional evolution through time. Importantly, it also demonstrates that phylogeny is a good predictor of trait variation in rubisco. Thus, as at other scales of biological organisation, the resemblance of rubisco physiochemical traits across species is shaped by the interconnected histories of these enzymes on the underlying hierarchical structure of the tree of life.

Given that the work in this thesis embodies one of few studies which investigate phylogenetic signal in enzymes, and as all of these studies have focused on rubisco, whether the findings presented herein apply to other enzymes in nature is not yet known. This uncertainty is particularly striking when considering the notable finding in Chapter 2 of this thesis which identifies a severe phylogenetic constraint on rubisco kinetic adaptation (Bouvier *et al.*, 2021). Moreover, this uncertainty is also further amplified when considering the finding in Chapter 3 of this thesis that demonstrates rubisco is encoded by one of slowest evolving genes on Earth (Bouvier, Emms and Kelly, 2024). On account of these two phenomena, rubisco is thus a particularly unusual enzyme that suffers from a very low rate of molecular and phenotypic diversity. Given that both of these features are central determinants that affect phylogenetic signal strength, a key goal of future research will be to investigate whether the presence of phylogenetic signal in rubisco is representative of a more general trend among other enzymes in nature which are not subject to the same severe constraints on their evolvability.

To address the above open-ended questions raised from this thesis, a valuable direction for future research would be to conduct a global phylogenetic signal analysis of the several thousand enzymes for which both sequence and kinetic data exist in BRENDA (Braunschweig Enzyme Database) (Schomburg, Chang and Schomburg, 2002; Schomburg *et al.*, 2017). Importantly, data gained from this work would allow a critical assessment of the dynamics shaping the evolution of all enzymes that are characterised in all taxa across the tree of life to evaluate the pervasiveness of phylogenetic signal. Following the lessons learned in Chapter 2 and Appendix A of this thesis (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010; Flamholz *et al.*, 2019; Bouvier *et al.*, 2021; Bouvier and Kelly, 2023), these data should also be used to promote the importance of phylogeny-aware approaches in comparative biology to avoid the “serious statistical problem” caused by phylogenetically non-independent observations (Felsenstein, 1985). Finally, these data will also prove instrumental in applied biology contexts to assess the extent by which phylogeny provides a comparative framework to extrapolate enzyme function.

Ultimately, the phylogenetically resolved interrogation of rubisco presented in this thesis has provided valuable new insight on this enzyme’s evolution, adaptation, and constraint (Bouvier *et al.*,

2021; Bouvier and Kelly, 2023; Bouvier, Emms and Kelly, 2024). As set by the precedent herein, continuing to embrace this 'phylogenetic revolution' to study enzyme function holds immense potential to advance diverse biological disciplines including evolution, biochemistry, biotechnology, and beyond.

5.3 The evolving landscape of rubisco: a revised biochemical–molecular constraint model helps to explain enzyme adaptation

5.3.1 Early evidence of constraints on rubisco adaptation

The kinetic traits of rubisco which determine the enzyme's reaction efficiency have been experimentally characterised for over four decades (Badger and Lorimer, 1976; Chollet and Anderson, 1976; Badger and Collatz, 1977; Jordan and Ogren, 1981, 1983; Yeoh, Badger and Watson, 1981; Seemann, Badger and Berry, 1984; Parry, Keys and Gutteridge, 1989). In brief, kinetic traits of interest which are experimentally accessible to biochemists include CO_2/O_2 specificity ($S_{C/O}$), CO_2 and O_2 turnover (k_{catC} and k_{catO}), the Michaelis constants for both CO_2 and O_2 substrates (K_C and K_O), the Michaelis constant for CO_2 in the presence of 21% O_2 (K_C^{air}), and the overall efficiencies of both CO_2 and O_2 reactions (k_{catC}/K_C and k_{catO}/K_O). Since the beginning of enzyme characterisation efforts, initial trends in rubisco natural variation indicated that species with high $S_{C/O}$ tended to exhibit low k_{catC} and *vice versa*. These observations resulted in scientific debate as to whether the oxygenation reaction of rubisco was essential, unavoidable, neither, or both (Lorimer and Andrews, 1973; Andrews and Lorimer, 1978). It also led to early speculation that the carboxylase and oxygenase functions of rubisco might inherently be linked (Badger and Lorimer, 1976; Chollet and Anderson, 1976), without formalising an underlying mechanism.

5.3.2 The seminal biochemical trade-off model posits a central compromise between specificity and CO_2 turnover

Investigating these early observations, Tcherkez, Farquhar, and Andrews performed a systematic cross-species analysis of the relationships between rubisco kinetic traits (Tcherkez, Farquhar and Andrews, 2006). Specifically, these authors interrogated the interspecific variation in kinetic trait and isotope-fractionation data across 19 phylogenetically diverse rubiscos isolated from bacteria, cyanobacteria, green algae, non-green algae, and both C_3 and C_4 plants. This analysis revealed that all kinetic traits were strongly correlated with $S_{C/O}$ across species, with the strongest dependency found between $S_{C/O}$ and k_{catC} (Tcherkez, Farquhar and Andrews, 2006). In addition to kinetic antagonism, dependencies were also reported between $S_{C/O}$ and both the Arrhenius activation energy of k_{catC}, as well as between $S_{C/O}$ and $^{13}C/^{12}C$ (but not $^{18}O/^{16}O$) isotope fractionation data (Tcherkez, Farquhar and Andrews, 2006).

On the basis of their findings, these authors proposed the first trade-off model to explain "the rubisco paradox" and the enzyme's apparent catalytic inefficiency (Tcherkez, Farquhar and Andrews, 2006). Specifically, it was hypothesised that the central compromise between $S_{C/O}$ and k_{catC} was grounded in inherent constraints underpinned by the rubisco reaction mechanism. More precisely, as selectivity between CO_2 and O_2 substrates is likely determined in the transition state of these reactions (i.e., the developing carboxyketone or peroxyketone, respectively), it was argued that selection for increased $S_{C/O}$ would drive the transition state of the carboxylation reaction to more closely resemble its carboxyketone product. In brief, this would result in an improvement in $S_{C/O}$ because the carboxylation transition state would now be maximally discriminable from the competing oxygenation transition state (thereby promoting CO_2 rather than O_2 attack). However, this structural change would also have the concomitant effect of causing a reduction in k_{catC} because the interactions required to stabilise the initial transition state prior to CO_2 addition would now also act to stabilise the similar structure of the carboxyketone product after it is formed (thereby slowing its downstream processing and the overall rate of reaction). In this way, this mechanism additionally explained the enhanced Arrhenius activation energy of the CO_2 reaction (due to an increased energy investment now required to process the carboxyketone product), as well as the differences in isotope fractionation of $^{13}C/^{12}C$

but not $^{18}O/^{16}O$ (with the structure of the latter transition state being wholly unaffected). Simply put, this theory thus postulates that rubisco cannot optimize $S_{C/O}$ and k_{catC} simultaneously, with the chemical origin of this trade-off underpinned by a compromise between either a more reactant-like (high k_{catC}, low $S_{C/O}$) or product-like (low k_{catC}, high $S_{C/O}$) carboxylation transition state (Tcherkez, Farquhar and Andrews, 2006).

Moving one step further, Tcherkez, Farquhar, and Andrews proposed that all extant rubisco were near-perfectly optimised to their prevailing CO_2, O_2, and temperature conditions (Tcherkez, Farquhar and Andrews, 2006). Specifically, this belief was forwarded based upon two distinct observations. First, species which maintain supra-atmospheric CO_2 concentrations around rubisco through operation of either a biochemical or biophysical CO_2 concentrating mechanism (CCM) exhibited faster k_{catC} and lower $S_{C/O}$ in comparison to their non-CCM counterparts (Tcherkez, Farquhar and Andrews, 2006). Thus, the position of an enzyme along the $S_{C/O}$~k_{catC} trade-off is associated with the prevailing photosynthetic conditions to which it was adapted. Second, the residual errors of all species data points from the line of best fit modelled to describe the $S_{C/O}$~k_{catC} trade-off were small. Thus, there is minimal latent optimisation left to occur in rubisco kinetics in the context of the underlying $S_{C/O}$~k_{catC} trade-off. To summarise, this study proposed that the catalytic mechanism of rubisco manifests in a reciprocal relationship between $S_{C/O}$ and k_{catC}, and that all extant rubisco are near perfectly optimised in relation to this trade-off in the context of their prevailing photosynthetic conditions.

5.3.3 A revised trade-off model proposes all kinetic traits are antagonistically linked to form a one-dimensional landscape

After the above seminal trade-off model was advanced, the nature of the physiochemical constraints shaping rubisco adaptation was later modified. Specifically, Savir and colleagues performed a re-examination of the kinetic relationships among 25 diverse Form I rubisco and instead reported that all kinetic traits were antagonistically linked with every other kinetic trait through simple power-law (i.e., log–log) relationships (Savir *et al.*, 2010). Whilst all of these kinetic trait dependencies were considerably strong, the most severe of these antagonisms was found between k_{catC} and K_C (Savir *et al.*, 2010).

Following observations of inexorable dependencies between all kinetic traits, Savir and colleagues proposed a new trade-off model. Specifically, it was argued that trade-offs between all combinations of enzyme kinetic traits manifest in a one-dimensional fitness landscape (Savir *et al.*, 2010). The implication of this landscape is such that the value of one kinetic trait determines that of all other traits. Or, to put it another way, a change in one kinetic trait leads to a proportional and inverse change in the values of all other kinetic traits. Thus, this proposed trade-off model is conceptually distinct from the seminal $S_{C/O}$~k_{catC} trade-off model, and instead postulates that interdependencies between all kinetic traits exist and act to impede overall enzyme improvement.

Given this updated understanding, these authors re-visited the question of rubisco optimality. Specifically, a modelling approach was used to compare kinetic trait values of sampled rubiscos with those estimated to be optimal within the context of the respective intracellular CO_2 and O_2 concentrations to which they are adapted. In agreement with the previous assertion made by (Tcherkez, Farquhar and Andrews, 2006), this analysis supported the belief that all extant rubisco were near-perfectly optimised in the context of the kinetic trait trade-offs that are purported to exist (despite differences in the nature of these trade-offs between these studies) (Savir *et al.*, 2010). Together, this led Savir and colleagues to propose that rubisco is perfectly optimised in a highly constrained one-dimensional fitness landscape. Thus, rubisco was posited to effectively be localised at a fitness peak in a landscape that is beset on all sides by trade-offs between its kinetic traits.

5.3.4 Biochemical trade-offs revisited in light of increased data availability

Since the above two seminal trade-off models were first formulated (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010), the amount of rubisco kinetic data across species has considerably increased owing to dedicated enzyme characterisation efforts (von Caemmerer *et al.*, 1994; Zhu *et al.*, 1998; Kubien *et al.*, 2008; Galmés *et al.*, 2014; Occhialini *et al.*, 2016; Orr *et al.*, 2016; Prins *et al.*, 2016; Sharwood *et al.*, 2016). Consequently, these expanded datasets have now enabled more detailed investigations into enzyme adaptation and constraint to a degree that was previously unattainable in earlier studies.

Taking advantage of increases in data availability, Flamholz and colleagues re-investigated the extent and magnitude of trade-offs between rubisco kinetic traits across an extended dataset of ~300

species (Flamholz *et al.*, 2019). This revised analysis found that the vast majority of all kinetic trait trade-offs were considerably attenuated compared to previous estimates (Flamholz *et al.*, 2019). Of particular importance, considerable relaxation was reported in the antagonistic relationship between $S_{C/O}$ and k_{catC} (as underpins the trade-off model proposed by (Tcherkez, Farquhar and Andrews, 2006)), as well as in the antagonistic relationship between k_{catC} and K_C (as is the central and most severe trade-off in the one dimensional fitness landscape model proposed by (Savir *et al.*, 2010)). In support of this conclusion, both Cummins (Cummins, 2021) and Iñiguez and colleagues (Iñiguez *et al.*, 2020) also reported similar weakened kinetic trait trade-offs among large datasets of sampled higher plant rubisco and sampled Form I rubisco, respectively.

Together, these revised analyses therefore strongly suggested that rubisco kinetic traits are less strictly coupled than first asserted (Flamholz *et al.*, 2019; Iñiguez *et al.*, 2020; Cummins, 2021). Importantly, they also begin to question the validity of each seminal trade-off model. Rubisco does not appear to be constrained by a central trade-off between $S_{C/O}$ and k_{catC} based on the mechanism of its reaction (Tcherkez, Farquhar and Andrews, 2006). Moreover, rubisco does not appear to exist in a one-dimensional fitness landscape underpinned by inextricable antagonistic dependencies between all combinations of its kinetic traits (Savir *et al.*, 2010).

5.3.5 Biochemical trade-offs revisited in light of phylogenetic signal

Alongside the reported attenuation of trade-offs in larger rubisco datasets, several small cross-species analyses found that rubisco kinetic trait relationships differed when applying phylogeny-aware comparative methods as compared to phylogeny-unaware comparative methods (Kubien *et al.*, 2008; Galmés *et al.*, 2014; Hermida-Carrera, Kapralov and Galmés, 2016; Orr *et al.*, 2016). As described in depth in Chapter 2 and Appendix A, this phenomenon arises because conventional phylogeny-unaware methods incorrectly assume that datapoints are distributed independent of phylogeny, creating a “serious statistical problem” that can lead to spurious conclusions (Felsenstein, 1985). In contrast, phylogeny-aware comparative methods provide a more robust framework to interrogate causal relationships in biology by correcting for the underlying relationships of species on the hierarchical tree of life (Felsenstein, 1985). In this way, the above phylogenetically controlled analyses of rubisco kinetic correlations are particularly important. This is because they suggest that

previous rubisco trade-off models inferred from conventional statistical methods may have been spuriously inferred due to the presence of phylogenetic non-independence in the data.

Building on the above results, in Chapter 2 I performed an extensive phylogenetically-resolved investigation into the evolutionary and biochemical factors constraining rubisco kinetic adaptation across a large dataset of species (Bouvier *et al.*, 2021). This work demonstrated that all rubisco kinetic traits exhibit a significant phylogenetic signal (Bouvier *et al.*, 2021). The presence of this signal is a key result because it means that trait variation across species is strongly reflective of phylogeny, such that closely related species exhibit more similar trait values than distantly related species (i.e., trait values are phylogenetically non-independent) (Bouvier *et al.*, 2021). Thus, this phylogenetic signal confirms fundamental expectations from evolution by natural selection. Importantly, it also shows that phylogenetic comparative tools must be used to correct for this signal when evaluating rubisco biochemical trade-offs.

In Chapter 2, I next applied phylogenetic comparative methods to infer the true extent of kinetic trade-offs in the absence of either phylogenetic bias or potential limitations in power caused by small sample size. This work demonstrated that apparent kinetic trait dependencies were considerably weakened compared to both the seminal trade-off studies (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010) as well as the more recent analyses of larger species datasets (Flamholz *et al.*, 2019; Iñiguez *et al.*, 2020; Cummins, 2021) (table 5.1). Of significance, these trade-offs were also weakened in comparison to correlation coefficients previously reported in another earlier and smaller phylogenetic analysis of rubisco kinetics (Orr *et al.,* 2016) (table 5.1). In fact when correctly accounting for phylogenetic signal in this thesis, I showed that only moderate trade-offs between both k_{catC} and K_C (and K_C^{air}), and between K_C and K_O, were robust among C_3 angiosperms (Bouvier *et al.*, 2021) (table 5.1). I also found that additional correlations between $S_{C/O}$ and k_{catC} and between $S_{C/O}$ and K_C (and K_C^{air}) became significant when the analysis was expanded to include all angiosperms and all photosynthetic organisms, but that these trait dependencies were respectively weaker in magnitude ($R^2 \leq 8.1$ %; table 5.1) (Bouvier *et al.*, 2021). In combination, this thesis thus contributes key insights toward our understanding of rubisco functional evolution. It demonstrates

Table 5.1. Correlation coefficient (R^2 values) of relationships between kinetic traits reported at 25 °C from Flamholz and colleagues (Flamholz *et al.*, 2019) (non-phylogenetic analysis), Orr and colleagues (Orr *et al.,* 2016) (phylogenetic analysis), and in Chapter 3 of this thesis (Bouvier *et al.*, 2021) (phylogenetic analysis). Missing entries are noted with N/A.

		Kinetic Dependency (% Variation Explained)							
	Taxa Sampled	$S_{C/O}$ ~ k_{catC}	$S_{C/O}$ ~ K_C	$S_{C/O}$ ~ K_C^{air}	$S_{C/O}$ ~ K_O	k_{catC} ~ K_C	k_{catC} ~ K_C^{air}	k_{catC} ~ K_O	K_C ~ K_O
(Flamholz *et al.*, 2019)	Form I	31.4	43.6	*N/A*	0	23.0	*N/A*	0	31.4
(Orr et al., 2016)	Angiosperm	88.4	60.2	58.5	4.1	52.4	45.3	18.2	9.1
(Bouvier *et al.*, 2021)	C_3 Angiosperm	0	0	0	1.1	37.4	34.9	2.2	11.6
	Angiosperm	8.1	6.2	5.3	0	30.9	31.6	0	8.7
	Form I	2.2	6.0	7.1	0	21.1	23.3	0	18.6

that biochemical trade-offs exist and constrain rubisco adaptation, but that these are less severe than initially thought when correcting for the presence of phylogenetic signal. Therefore, the catalytic mechanism of rubisco has allowed for the vast majority of all rubisco kinetic traits to evolve largely independently of one another.

5.3.6 Rubisco phylogenetic signal – questions and advancements

After Chapter 2 was published in a peer-reviewed journal (Bouvier *et al.*, 2021), this work raised interest from the wider community. Specifically, Tcherkez and Farquhar argued in an Opinion article that the phylogenetic signal in rubisco kinetics might be an artefact caused by several systematic biases associated with the analysis of closely related species, imbalanced sampling across the tree of life, the use of the *rbcL* phylogenetic tree to infer species relationships, and both laboratory-to-laboratory and methodology-to-methodology variability in trait measurements (Tcherkez and Farquhar, 2021). Together, this led these authors to arrive at the conclusion that the phylogenetic signal discovered in rubisco kinetics is primarily driven by differences in the photosynthetic conditions to which these rubiscos are respectively adapted to (i.e., C_3 versus C_4 photosynthesis in plants) rather than being reflective of phylogeny.

In Appendix A, I critically analysed each of the above claims that were made and showed that these were unfounded (Bouvier and Kelly, 2023). Specifically, I demonstrated that a strong and pervasive phylogenetic signal is consistently reproduced in rubisco across an extended range of additional analyses designed to control for each systematic bias that was purported to exist. Thus, in support of the conclusions in Chapter 2, this work shows that the presence of phylogenetic signal in rubisco kinetics is genuine and must be properly accounted for when analyzing rubisco kinetic trait trade-offs.

5.3.7 Rubisco kinetic adaptations associated with C_3 versus C_4 photosynthesis

Whilst the interaction with Tcherkez and Farquhar diverted some momentum from the forward progress of this thesis, it facilitated a valuable platform for extended discussion. Specifically, this exchange enabled me to provide additional clarification on the concepts of phylogenetic non-independence, the relevance of this phenomenon to comparative biology, and how this statistical issue may be corrected for using phylogenetic comparative tools (Bouvier and Kelly, 2023). I am hopeful that this response and extended set of analyses will help highlight the importance of phylogenetically-controlled comparative analyses and advocate for their more widespread utilization in biochemistry and other molecular disciplines (see Chapter 5.2).

Importantly, Appendix A also offered an opportunity to critically review the assertion made by Tcherkez and Farquhar that rubisco kinetic variation in nature is primarily shaped by C_3 versus C_4 photosynthetic differences among species. Indeed, this is a potentially confusing issue. To explain, the phylogenetically-resolved analysis in Chapter 2 demonstrates that adaptive changes in rubisco kinetics occur between C_3 and C_4 plants (Bouvier *et al.*, 2021). Specifically, rubisco encoded by C_4 plants are adapted to be faster (higher k_{catC}) but less specific for CO_2 (lower $S_{C/O}$, higher K_C) compared to their C_3 counterparts (Bouvier *et al.*, 2021). These same observations have been reported previously between C_3 and C_4 plants (Yeoh, Badger and Watson, 1980, 1981; Seemann, Badger and Berry, 1984; Wessinger, Edwards and Ku, 1989; Sage, 2002; Kubien *et al.*, 2008; Iñiguez *et al.*, 2020; Cummins, 2021), and are indicative of broader trends that exist between CCM and non-CCM containing species across the tree of life (Badger, Hanson and Price, 2002; Young *et al.*, 2016; Young and Hopkinson, 2017; Goudet *et al.*, 2020; Iñiguez *et al.*, 2020; Capó-Bauçà,

Iñiguez and Galmés, 2024). Thus, in agreement with initial suggestions (Lorimer and Andrews, 1973; Andrews and Lorimer, 1978), and those of Tcherkez and Farquhar (Tcherkez, Farquhar and Andrews, 2006; Tcherkez and Farquhar, 2021), rubisco kinetic adaptation is shaped by selective pressures determined by biochemical niche.

Whilst the above is true, it is incorrect to state that the adaptive differences between C_3 and C_4 plants are responsible for the presence of phylogenetic signal in rubisco kinetics. Indeed, in both Chapter 2 and Appendix A, I demonstrate that a strong phylogenetic signal in rubisco kinetics is pervasive at all scales across the tree of life including among Form I rubisco, among all angiosperms, and importantly, also when the analysis is restricted to only C_3 angiosperms (Bouvier *et al.*, 2021; Bouvier and Kelly, 2023). In fact, the presence of this phylogenetic signal is stronger among C_3 species compared to analyses where all angiosperms or all Form I rubisco are studied together (Bouvier *et al.*, 2021; Bouvier and Kelly, 2023). The phylogenetic signal found can therefore not be attributed to photosynthetic differences. This result is concordant with evolutionary and statistical expectations because C_4 photosynthesis is a convergent adaptation that has occurred at disparate positions along the phylogenetic tree. As such, inclusion of C_4 species in the analysis would degrade the statistical co-variation between the kinetic traits and phylogenetic distances of the sampled rubisco (the precise feature measured by phylogenetic signal). Thus, rubisco kinetics have co-evolved with CCMs as posited by Tcherkez and Farquhar (Tcherkez and Farquhar, 2021) and as demonstrated in Chapter 2 (Bouvier *et al.*, 2021), but this is not a driver of the phylogenetic signal in rubisco kinetics. In fact, this phenomenon has the opposite effect of weakening the strength of this signal.

5.3.8 The relevance of C_3 versus C_4 kinetic adaptations and biochemical trade-offs

Together, the results from this thesis and other recent studies reveal that kinetic trait trade-offs are attenuated in light of both increases in rubisco kinetic data (Flamholz *et al.*, 2019; Iñiguez *et al.*, 2020; Cummins, 2021) as well after correction of phylogenetic signal in this data (Kubien *et al.*, 2008; Galmés *et al.*, 2014; Hermida-Carrera, Kapralov and Galmés, 2016; Orr *et al.*, 2016; Bouvier *et al.*, 2021; Bouvier and Kelly, 2023). However, whilst a statistical explanation for this attenuation related to phylogenetic correction has been provided in Chapter 2 and Appendix A, the impact of sample size on this attenuation has yet to be fully addressed.

Indeed, both seminal rubisco kinetic trade-off studies were based on small datasets of species (19; (Tcherkez, Farquhar and Andrews, 2006) and 25; (Savir *et al.*, 2010)) that likely suffered issues with statistical power. However, these analyses also sampled autotrophs from phylogenetically diverse lineages and disparate photosynthetic contexts including carboxysome-containing cyanobacteria, pyrenoid-containing green and non-green algae, and both C_3 and C_4 plants (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010). In comparison, all later analyses were either exclusively composed of, or largely dominated by, C_3 angiosperms (Flamholz *et al.*, 2019; Iñiguez *et al.*, 2020; Bouvier *et al.*, 2021; Cummins, 2021; Bouvier and Kelly, 2023). As there exists a strong photosynthetic dependency in rubisco kinetics (Yeoh, Badger and Watson, 1980, 1981; Seemann, Badger and Berry, 1984; Wessinger, Edwards and Ku, 1989; Badger, Hanson and Price, 2002; Sage, 2002; Kubien *et al.*, 2008; Young *et al.*, 2016; Young and Hopkinson, 2017; Goudet *et al.*, 2020; Iñiguez *et al.*, 2020; Bouvier *et al.*, 2021; Cummins, 2021; Capó-Bauçà, Iñiguez and Galmés, 2024), it is possible that this has helped drive artefactual correlations in earlier studies compared to later studies owing to differences in dataset size and composition.

I postulate here that the Simpson's paradox may provide a plausible explanation for how photosynthetic dependencies in the data could help further drive artefactual correlations (Blyth, 1972). In brief, the Simpson's paradox arises when combining different subpopulations (i.e., rubisco adapted to different photosynthetic contexts) in analyses produces misleading relationships that may reverse or disappear when those subgroups are analysed separately. Indeed, strong evidence for the presence of this statistical phenomenon is provided in Chapter 2, in which a significant correlation between $S_{C/O}$ and k_{catC} is produced in the analysis of all angiosperms and all Form I rubisco but not when the analysis is restricted to C_3 angiosperms (table 5.1) (Bouvier *et al.*, 2021). Together, this result suggests that the purported association between $S_{C/O}$ and k_{catC} is not a strict causal relationship (which would otherwise be expected to occur at all levels of biological organisation, including among C_3 angiosperms). Instead, this trade-off more likely reflects an artefact caused by photosynthetic dependencies in the rubisco kinetic data underpinned by differences in selection pressures on enzyme kinetic trait evolution. As the extent of rubisco kinetic adaptation to a CCM is known to vary with its efficiency and age across organisms (Meyer and Griffiths, 2013), this effect would be further

exaggerated among analyses of small species datasets sampled from phylogenetically diverse lineages and widely disparate photosynthetic contexts (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010). Thus, the photosynthetic dependency in rubisco kinetics is not responsible for phylogenetic signal. However, it could play a role in helping to explain contradictions in kinetic trait associations between studies in conjunction with phylogenetic signal.

5.3.9 Why do rubisco kinetic trait differences exist between C_3 versus C_4 photosynthesis if not for biochemical trade-offs?

The above extended discussion serves to highlight the importance of considering cause and effect. In the seminal $S_{C/O}$ ~ k_{catC} trade-off model (Tcherkez, Farquhar and Andrews, 2006), differences in $S_{C/O}$ and k_{catC} values between CCM and non-CCM containing species are postulated to be caused by an underlying biochemical trade-off between these traits. However, I argue above that it is more likely that differences in kinetic adaptation cause an artefactual correlation between these traits (the inverse mechanism). This led me to postulate that differences in correlations between studies may arise from phylogenetic signal in addition to photosynthetic dependencies in the data.

To avoid confusion, it is worth pausing here to assess how these photosynthetic dependencies in $S_{C/O}$ and k_{catC} may arise in the absence of trade-off. As described extensively in Chapter 2 and Chapter 3, differences in rubisco kinetic traits between C_3 and C_4 species align with the differences in selection pressures in each photosynthetic context (Bouvier *et al.*, 2021; Bouvier, Emms and Kelly, 2024). Within a C_3 leaf, $S_{C/O}$ and k_{catC}/K_C are the major kinetic determinants of photosynthetic efficiency (Farquhar, von Caemmerer and Berry, 1980). This makes sense when considering the low concentration of CO_2 in the C_3 mesophyll chloroplast (~8 μM) and the declining ratio of atmospheric CO_2:O_2 through geological time (Figure 3.1). This expectation also aligns with the results in Chapter 3 in which I find that rubisco in C_3 angiosperms experience continual directional selection for enhanced $S_{C/O}$, k_{catC}, and k_{catC}/K_C (Bouvier, Emms and Kelly, 2024).

In contrast to the biochemical niche of rubisco in a C_3 leaf, transition toward C_4 photosynthesis is associated with different selective pressures. Specifically, rubisco becomes exposed to supra-atmospheric concentrations of CO_2 (~160 μM) which saturate the enzyme's active site and suppress its oxygenase activity (Jenkins, Furbank and Hatch, 1989). In this scenario, k_{catC} would be expected

to be under positive selection to maximise the frequency of carboxylation reactions, whilst $S_{C/O}$, K_C and K_O would all be expected to evolve neutrally in the absence of selection. Indeed, this aligns with empirical observations in which k_{catC} is found to be the sole rubisco kinetic determinant of photosynthetic efficiency in C_4 plants (Badger and Andrews, 1987; Sage, 2002; Sharwood, 2017). Together, the above differences in selection pressures explain the findings in Chapter 2 (namely, that rubisco in C_3 species exhibit improved $S_{C/O}$ but reduced k_{catC}, and that rubisco in C_4 plants exhibit improved k_{catC} but reduced $S_{C/O}$) in the absence of a strong trade-off between these traits. In doing so, it helps to explain how this natural variation could then be mistaken for kinetic trait trade-offs in downstream analyses.

5.3.10 The discovery of a phylogenetic constraint

Given the discovery of weakened kinetic trait trade-offs in Chapter 2 (Bouvier *et al.*, 2021), I posited in this thesis that biochemical constraints are insufficient to solely explain the rubisco paradox. Accordingly, in Chapter 2 I explored alternate mechanisms that may also help drive the enzyme's purported inefficiency. In doing so, I discovered that phylogenetic constraints explained more variation in rubisco kinetic traits than the combined action of all kinetic trait trade-offs (Bouvier *et al.*, 2021). In this context, phylogenetic constraint quantifies how much variance in trait data can be explained by phylogeny, and is broadly interpreted to measure an inherent limitation on functional adaptation (Ives, 2019). In summary, Chapter 2 thus arrives at the conclusion that phylogenetic constraints co-limit rubisco adaptation in combination with the canonical kinetic trait trade-offs between k_{catC} and K_C and between K_C and K_O.

5.3.11 Phylogenetic constraints and the prevailing question of optimality

Beyond the key findings in Chapter 2, several important questions are left unanswered from this work. Notably, whilst a strong phylogenetic constraint is discovered to limit rubisco kinetic adaptation, the underlying mechanism responsible for this phenomenon has never been formally resolved. This gap in our knowledge means that the exact biological basis of the rubisco paradox is still not precisely understood. Moreover, it also remains unclear whether the attenuation of kinetic trait trade-offs and the discovery of phylogenetic constraints challenge the paradigm that rubisco is "near-perfectly optimised". For instance, whilst both seminal trade-off models suggest that rubisco is optimally

adapted (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010), this argument is predicated upon the context of the exaggerated trade-offs that were reported. Given that these trade-offs are now widely appreciated to be weaker, it accordingly motivates a re-evaluation of enzyme optimality.

Because phylogenetic constraints present an important limitation on rubisco adaptation (and optimisation), I speculated in this thesis that resolving the question of rubisco optimality may help provide insight into the phylogenetic constraint identified. Following this lead, in Chapter 3 I sought to formally investigate rubisco optimality using an evolutionary framework. Specifically, I performed a phylogenetically-resolved analysis to quantify the pace of rubisco evolution to assess whether the enzyme is in evolutionary stasis (consistent with being optimally adapted) or whether it has been adaptively evolving for improved function (consistent with being non-optimally adapted). The first notable finding from this work revealed that *rbcL,* which encodes the catalytic large subunit of rubisco, is among the 1% of slowest evolving genes and enzyme-encoding genes on Earth (Bouvier, Emms and Kelly, 2024). The second notable finding from this work revealed that despite slow sequence evolution, rubisco catalysis and CO_2 assimilation capacity has been continually improving during the diversification of C_3 plants (Bouvier, Emms and Kelly, 2024). This latter result is in alignment with other studies which demonstrate a 2-fold improvement in rubisco carboxylation efficiency over the course of land plant evolution from bryophytes to flowering plants (Galmés *et al.,* 2014). All of these results are also consistent with the discovery of widespread positive selection in *rbcL* (Kapralov and Filatov, 2007; Galmés *et al.,* 2014). Together, this work thus provide a strong basis to argue that rubisco is not locked in stasis, but is instead adaptively evolving for improved efficiency. Comparable to the delayed discovery of plate tectonics, a slow rate of change in rubisco has been widely mistaken for stasis and a perfectly evolved enzyme.

5.3.12 Phylogenetic constraints, molecular constraints, and the rubisco paradox

Aside from shedding new insight on rubisco optimality, I argue here that the improved understanding of rubisco kinetic evolution provided in Chapter 3 is of key relevance to the wider discussion of phylogenetic constraint. Considering that (1) rubisco molecular evolution is exceedingly slow, and (2) rubisco catalysis and CO_2 assimilation capacity improve at a rate which is dependent on the extent of RbcL molecular evolution across species, it is reasonable to speculate that slow molecular

evolution presents a significant limitation on rubisco functional adaptation. For instance, as plant rubisco accumulate only one nucleotide substitution every 0.9 million years (Bouvier, Emms and Kelly, 2024), most amino acid changes would be restricted to those within one nucleotide change from the existing sequence. Similarly, as plant rubisco accumulate one amino acid substitution on average every 7.2 million years (Bouvier, Emms and Kelly, 2024), two or more amino acid changes are not likely to arise simultaneously for enzyme improvement. Because sequence determines kinetics, these combined limitations would thus restrict the adaptive potential of rubisco by reducing the landscape accessible for its improvement.

The above limitation, which I term "the molecular constraint hypothesis", offers a plausible explanation that helps to account for why rubisco kinetics appear better suited to the high CO_2 and low O_2 conditions under which the enzyme evolved, rather than the low CO_2 and high O_2 conditions of the present day (Figure 3.1). This theory also explains the extremely low genetic and kinetic variation among diverse rubiscos in nature (Kellogg and Juliano, 1997; Flamholz *et al.*, 2019; Bouvier *et al.*, 2021; Bouvier and Kelly, 2023; Bouvier, Emms and Kelly, 2024). Importantly, molecular constraint also offers a hypothesis to understand the mechanism of the severe phylogenetic constraint on rubisco kinetic adaptation reported in Chapter 2 (Bouvier *et al.*, 2021). In doing so, the molecular constraint hypothesis thus unifies disparate aspects of our understanding of rubisco and offers a new perspective to help appreciate the rubisco paradox.

5.3.13 Future work needed to resolve molecular constraints

Although the molecular constraint hypothesis predicts that slow sequence evolution functionally constrains enzyme adaptation, this slow sequence evolution is not in itself the biological mechanism that is responsible for the phylogenetic constraint. Instead, the responsible mechanism is the underlying factors which cause the slow molecular evolution to arise in the first place. As discussed in Chapter 2 and Chapter 3, the biological basis of this slow molecular evolution (and phylogenetic constraint) is likely caused by a multitude of factors which are specific to rubisco and include its high transcript and protein abundance, constraints imposed by selection to preserve function (including stability–activity trade-offs and kinetic trait trade-offs), and constraints imposed by its intermolecular

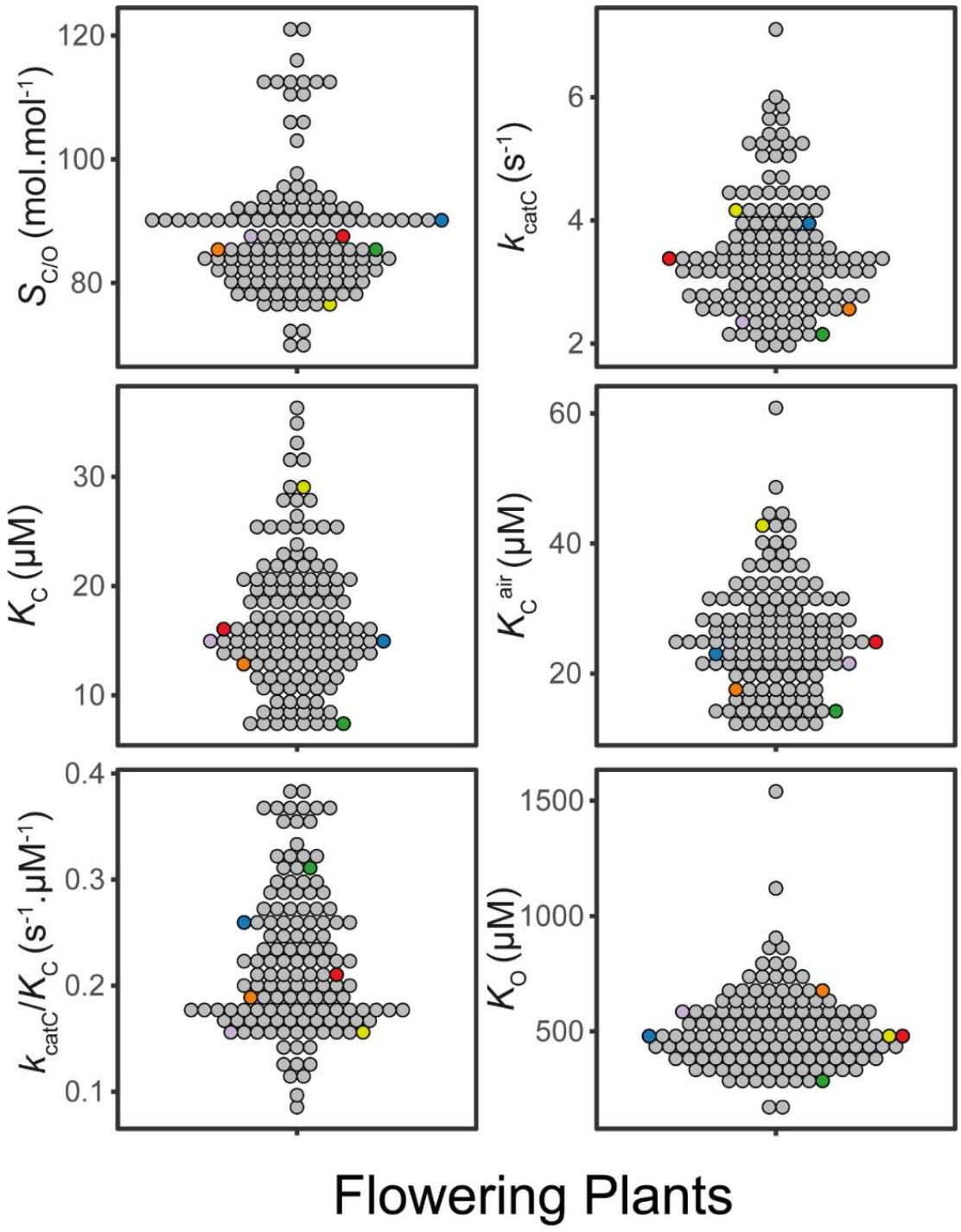


Figure 5.2. The distributions of values for rubisco kinetic traits in angiosperms. Key crops of agricultural interest are highlighted in different colours. Green: *Oryza sativa* subsp. *indica* (rice; C_3, monocot). Red: *Triticum aestivum* (wheat; C_3, monocot). Blue: *Hordeum vulgare* (barley; C_3, monocot). Yellow: *Zea mays* (maize; C_4, monocot). Orange: *Glycine max* (soybean; C_3, dicot). Lilac: *Manihot esculenta* (cassava; C_3, dicot). Kinetic trait units are shown on y axis labels.$S_{C/O}$: specificity. k_{catC}: carboxylase turnover per site. K_C: the Michaelis constant for CO_2. K_C^{air} the inferred Michaelis constant for CO_2 in 20.95% O_2 air. k_{catC}/K_C: carboxylation efficiency. K_O: the Michaelis constant for O_2. Data is analysed from (Bouvier *et al.*, 2021).

interactions with a large diversity of chaperones that are essential for its correct folding, assembly and metabolic regulation (Bouvier *et al.*, 2021; Bouvier, Emms and Kelly, 2024). Importantly, in plants and other eukaryotes, the evolvability of rubisco would also be constrained by it being encoded in the plastid genome which is inherited maternally in the absence of recombination (Wolfe, Li and Sharp, 1987; Birky, 2001; Smith, 2015). Indeed, a recent phylogenomic analysis has identified that three factors together explain ~50% of variation in molecular evolution rates across angiosperm chloroplast encoded genes that include the distance from the origin of replication, amino acid composition (affecting substitution tolerance), and mRNA abundance (Robbins and Kelly, 2023). This finding reveals that "much of the slow evolution rate" of rubisco can be explained by factors largely unrelated to its function, offering important insights regarding the nature of some molecular constraints affecting the enzyme's adaptive potential. Moving forward, an evaluation of the precise nature of the remaining determinants of rubisco's slow molecular evolution rate presents a key goal of ongoing work that may promise far-reaching implications for our understanding of both the enzyme's adaptation and for the rubisco paradox.

5.3.14 Potential implications for rubisco engineering

Improving plant rubisco has represented a long-standing goal of scientific research (Gutteridge *et al.*, 1984; Gutteridge, Lorimer and Pierce, 1988; Prywes *et al.*, 2022). However, progress associated with this challenge has to date been limited. Despite this frustration, I present several key findings in this thesis that suggest rubisco is a potentially viable engineering target. The increased plasticity in rubisco biochemical trade-offs demonstrate that the enzyme's kinetic traits can be manipulated largely independently of one another for improvement. Moreover, it should also be possible to bypass phylogenetic (molecular) constraints on rubisco kinetic adaptation through dedicated engineering. Even under the unlikely scenario that rubisco is near-perfectly optimised (as I argue against here), the enzyme variants encoded by many of our major crops do not represent the pinnacle of efficiency across the plant kingdom (Figure 5.2). Thus, there is latent potential to target rubisco for crop improvement.

Harnessing this optimism, scientists should not lose hope for rubisco. Instead, we should embrace all tools available in the current synthetic and computational biology revolutions to continue working

toward the goal of a better enzyme. Laboratory-based directed evolution platforms offer an inexpensive and high-throughput method to screen for improved variants and have shown promising success in Form II rubisco and non-plant Form I rubisco (Wilson, Alonso and Whitney, 2016; Wilson *et al.*, 2018; Zhou and Whitney, 2019; Gionfriddo, Rhodes and Whitney, 2023). Now, recent achievements in assembling plant rubisco in bacteria (Aigner *et al.*, 2017; Conlan and Whitney, 2018; Lin *et al.*, 2022; Sally Buck *et al.*, 2023) offer the exciting prospect for these technologies to be applied to improve crop variants. Complementing these powerful laboratory-based screening platforms, phylogenetic analyses of sequence data have long offered a powerful method to map rubisco sites of functional importance as well as to identify target mutations for improvement (Kapralov and Filatov, 2007; Galmés *et al.,* 2014). More recently, this capability has been further extended by the development of machine learning pipelines which are able of accurately predicting rubisco kinetic trait values from input sequence data (Iqbal, Lisitsa and Kapralov, 2022). When also coupled alongside recent advances in computational protein structure prediction (Jumper *et al.*, 2021), it is evident that these combined break-throughs will pave the way for a new era of *in silico* screening and computational-aided enzyme design. Whereas previous engineering approaches have largely focused on testing single mutations through a trial-and-error approach, these emerging laboratory- and computational-based technologies will offer unprecedented capacity to unravel the rubisco sequence–function landscape and systematically explore the vast potential for its optimization. The long-standing goal of enhancing rubisco may, finally, be within reach.

5.4 Metabolic engineering of stomatal density presents a potential future target for crop improvement

The evolution of land plants marked an increase in the complexity of photosynthetic organisms. Associated with this transition, a new suite of physical resistances emerged that impede the supply of CO_2 substrate to the site of carbon fixation in the plant leaf (Evans *et al.*, 2009; Flexas *et al.*, 2013). As stomata overcome a key obstacle along this CO_2 diffusion path and serve as the central interface

for plant CO_2 and H_2O gas exchange, these pores have represented a key crop improvement target for over four decades (Farquhar and Richards, 1984).

In Chapter 4 of this thesis, I investigated whether altering cellular levels of nicotinamide adenine dinucleotide (NAD^+) could provide a novel metabolic handle for manipulating stomatal density. For this purpose, the model plant *Arabidopsis thaliana* was engineered to express a bacterial water-forming NAD(P)H oxidase (NOX) that was targeted to perturb $NAD(P)^+$/NAD(P)H status in undifferentiated stomatal precursor cells (Bouvier and Kelly, 2024). Specifically, NOX was selected for these experiments because it catalyses the oxidation of NAD(P)H to $NAD(P)^+$ in the inverse direction to that of the water-splitting reaction of photosystem II during oxygenic photosynthesis [$2NAD(P)H + 4H^+ + O_2 \rightarrow 2NADP^+ + 2H_2O$]. As such, NOX was hypothesized to provide a powerful genetically encoded tool for manipulating cellular redox status with minimal disruption to other metabolic processes.

Indeed, the experiments in Chapter 4 reveal that all NOX-expressing lines exhibit a ~20–30% reduction in measured abaxial stomatal densities (Bouvier and Kelly, 2024). Moreover, this reduced stomatal density phenotype was reproduced when NOX was targeted to the cytosol, chloroplast stroma, and chloroplast intermembrane space of undifferentiated epidermal cells, respectively (Bouvier and Kelly, 2024). Building upon key findings from earlier studies (de Souza Chaves *et al.*, 2019; Feitosa-Araujo *et al.*, 2020; Falquetto-Gomes *et al.*, 2024), this work thus represents an important advancement that further underscores the role of NAD^+ as an emerging negative regulator of stomatal patterning.

In addition to identifying NAD^+ as a novel metabolic engineering target, the results in Chapter 4 also highlight the potential utility of this proof-of-concept approach for crop improvement. This is because I discovered that all engineered plants exhibit a carbon-neutral enhancement in WUE, underpinned by a significant reduction in stomatal conductance whilst maintaining wild-type levels of net CO_2 assimilation (Bouvier and Kelly, 2024). Importantly, I also demonstrate that this improvement in plant photosynthesis-water relations translates to an enhancement in biomass accumulation under water-deficit conditions without any measured growth penalty under optimal watering conditions (Bouvier

and Kelly, 2024). Together, this genetic engineering method thus offers a promising strategy to enhance plant productivity and climate-resilience, with possible implications for agriculture.

Despite the potential promise of the genetic engineering technology in Chapter 4, it should be noted that the application of this approach in the context of crop improvement is at current limited. Specifically, this limitation stems from the finding that all NOX-expressing lines suffered both a decrease in overall seed yield as well as a reduction in leaf mass per area (LMA = [leaf mass (g) / leaf area (m^2)]) (Bouvier and Kelly, 2024). Whilst this alteration in LMA is an anticipated consequence of cross-talk in the developmental regulation of stomata and the leaf mesophyll (Dow, Berry and Bergmann, 2017; Lundgren *et al.*, 2019), the impaired seed yield is an unexpected, and evidently, an unfavourable result. This yield penalty is particularly surprising given that it was measured in plants under water-deficit experiments, which otherwise displayed enhanced vegetative growth as well as traits indicative of improved reproductive success (e.g., taller inflorescences) compared to wild-type controls (Bouvier and Kelly, 2024). Moreover, no physiological change associated with either NOX expression or the reduced stomatal density phenotype could be detected in the engineered plants to explain this yield penalty (Bouvier and Kelly, 2024). Accordingly, this led me to hypothesise in Chapter 4 that this result was attributable to off-target effects from the use of the chlorophyll a/b binding protein 3 (CAB3) promoter for driving transgene expression (Mitra, Choi and An, 1989; Klepikova *et al.*, 2016). To address this, a key goal for future research will be to investigate whether these yield penalties can be mitigated by using alternative promoters that offer improved spatial and temporal control of NOX activity.

Building upon this proposed improvement to the proof-of-concept approach herein, future research may also investigate whether an analogous stomatal density phenotype can be produced by overexpression of plant endogenous NADH-consuming enzymes. Indeed, the results generated in this thesis suggest that this goal should be feasible if $NAD(P)^+$/NAD(P)H ratios are sufficiently perturbed. If accomplished, substituting NOX for a plant native enzyme (under the regulation of plant native promoters and terminators) would offer the advantage of imposing less regulatory hurdles in a potential crop improvement application. Coupled with efforts to address the yield penalty above,

these future research directions hold the potential to help secure long-term translational impact from this thesis.

Beyond its applied significance, the results in Chapter 4 also have important implications for our fundamental understanding of stomatal physiology and evolution. As discussed above, all NOX-expressing lines characterised herein demonstrate that moderate reductions in stomatal density can occur without any penalty in CO_2 assimilation rate (Bouvier and Kelly, 2024). This in itself represents an interesting result, and aligns with similar discoveries in barley and rice mutants that also exhibit moderate reductions in stomatal densities (Hughes *et al.*, 2017; Caine *et al.*, 2019). Together, these collective studies suggest that there is a latent capacity to optimise the balance between carbon uptake and water loss across a wide range of plant and crop species. However, the underlying basis for why stomatal density, CO_2 assimilation, and transpiration are not tightly co-ordinated under current conditions is unclear. Although several hypotheses are possible (Harrison *et al.*, 2020), I propose one plausible explanation is that plant stomatal densities are not well-adapted to current CO_2 concentrations. For instance, although stomatal numbers have exhibited a general decline over the past 200 years in response to rising CO_2 levels (Woodward, 1987), the unprecedented rate of this CO_2 change over such a short evolutionary timeframe has likely surpassed the capacity of plants to become optimally adapted. Analogous to the scenario proposed for rubisco in Chapter 3, this hypothesis thus posits that stomatal densities in extant plants might be better suited to the lower atmospheric CO_2 levels of the past rather than the higher atmospheric CO_2 levels of the present-day. Importantly, this hypothesis would also predict an imbalance between the optimal rates of operational stomatal conductance and photosynthesis that are concordant with the effects described in Chapter 4 and other recent studies (Hughes *et al.*, 2017; Caine *et al.*, 2019). Together, this hypothesis thus serves as an interesting basis for discussion and future investigation.

Given that NAD^+ negatively regulates stomatal development, it is also tempting to speculate that this signalling molecule might be involved in stomatal density responses to both light and CO_2. For instance, stomatal density in newly emerging leaves is known to increase with light intensity, such that plants grown under high irradiance develop leaves with increased numbers of stomata compared to plants grown under lower irradiance (Vigna *et al.*, 1980; Tichá, 1982; James and Bell,

2000; Lake *et al.*, 2001). Moreover, stomatal density also inversely correlates with atmospheric CO_2, such that plants grown under high CO_2 concentrations produce leaves with reduced numbers of stomata compared to plants under lower CO_2 concentrations (Woodward, 1987; Woodward and Bazzaz, 1988; McElwain and Chaloner, 1995; Beerling, McElwain and Osborne, 1998; Beerling, Osborne and Chaloner, 2001; Woodward, Lake and Quick, 2002; Franks and Beerling, 2009). A plausible link between NAD^+ signalling and these stomatal responses to both light and CO_2 arise because the direction of these density changes correlate with anticipated changes in cellular redox status. Specifically, increasing illumination results in plant photosynthetic cells with lower ratios of intracellular cytosolic NAD^+/NADH (consistent with an increase in leaf stomatal density based on the expected direction of NAD^+-mediated regulation) (Lim *et al.*, 2020), whilst increasing CO_2 levels results in plant photosynthetic cells with higher ratios of intracellular cytosolic NAD^+/NADH (consistent with a decrease in leaf stomatal density based on the expected direction of NAD^+-mediated regulation) (Igamberdiev and Gardeström, 2003). Further supporting this hypothesis, ABA is known to be implicated in both CO_2- and light-induced stomatal density responses (Movahedi *et al.*, 2021), in agreement with the proposed ABA-mediated mechanism of action of NAD^+ regulation (de Souza Chaves *et al.*, 2019; Feitosa-Araujo *et al.*, 2020; Falquetto-Gomes *et al.*, 2024). Together, it is thus feasible that NAD^+ integrates environmental cues of CO_2 and light to mediate stomatal density responses to changes in prevailing growth conditions. Whilst speculation, the above hypothesis warrants further investigation as could mean that NAD^+ provides a crucial missing link which unifies our understanding of the metabolism and physiology of plants with their development and adaptation.

Ultimately, the work presented in Chapter 4 advances our understanding of stomatal biology and shines a light on the emerging role of NAD^+ as a novel regulator of stomatal development. In doing so, it also highlights the viability of targeting NAD^+ to manipulate stomatal patterning in synthetic biology applications, with potential implications for agriculture. Whilst the exact mechanism of NAD^+-mediated stomatal regulation and its adaptive significance are still yet to be determined, it is evident that further unravelling this interaction between metabolism and development may help unlock significant innovations at the frontier of basic and applied biology.

5.5 Concluding remarks

Photosynthesis is the engine of life. It shaped the conditions on Earth which enabled us to evolve, and it continues to sustain us by producing both the oxygen that we breathe as well as the food that we consume. At the heart of this process, carbon fixation converts atmospheric CO_2 into the sugars which underpin the global food chain. This thesis provides several key insights which advance our understanding of the past, present, and future potential of plant carbon fixation in nature as well as in applied engineering contexts. In light of the conclusions presented, I argue that our agricultural crops did not emerge on Earth in an optimised state, and they are not perfectly optimised now. However, evolution combined with millennia of plant breeding has provided us with a robust foundation for future advancements. Whilst new innovations in crop improvement are needed to help feed the world in the 21st century, these will not have to reinvent the wheel. Perhaps, just to realign it. Optimising plant photosynthesis and carbon fixation offers one potential avenue by which this might be achieved.

5.6 References

Abouheif, E. (1999) 'A method for testing the assumption of phylogenetic independence in comparative data', Evolutionary Ecology Research, 1(8), pp. 895–909.

Aigner, H. et al. (2017) 'Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2', Science, 358(6368), pp. 1272–1278. Available at: https://doi.org/10.1126/science.aap9221.

Ajawatanawong, P. (2017) 'Molecular Phylogenetics: Concepts for a Newcomer', in Advances in Biochemical Engineering/Biotechnology, pp. 185–196. Available at: https://doi.org/10.1007/10_2016_49.

Andrews, T.J. and Lorimer, G.H. (1978) 'Photorespiration - still unavoidable?', FEBS Letters, 90(1), pp. 1–9. Available at: https://doi.org/10.1016/0014-5793(78)80286-5.

APG (1998) An Ordinal Classification for the Families of Flowering Plants, Annals of the Missouri Botanical Garden. Available at: https://doi.org/10.2307/2992015.

APG (2016) 'An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV', Botanical Journal of the Linnean Society, 181(1), pp. 1–20. Available at: https://doi.org/10.1111/boj.12385.

Badger, M.R. and Andrews, T.J. (1987) 'CO-Evolution of Rubisco and CO2 Concentrating Mechanisms', Progress in Photosynthesis Research, pp. 601–609. Available at: https://doi.org/10.1007/978-94-017-0516-5_128.

Badger, M.R. and Collatz, G.J. (1977) 'Studies on the kinetic mecha- nism of ribulose-1,5-bisphosphate carboxylase and oxygenase reactions with particular reference to the effect of temperature on kinetic parameters', Carnegie Institution of Washington Yearbook, 76(1961), pp. 355–361. Available at: http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract.

Badger, M.R., Hanson, D. and Price, G.D. (2002) 'Evolution and diversity of CO2 concentrating mechanisms in cyanobacteria', Functional Plant Biology, 29(2–3), pp. 161–173. Available at: https://doi.org/10.1071/pp01213.

Badger, M.R. and Lorimer, G.H. (1976) 'Activation of ribulose-1,5-bisphosphate oxygenase. The role of Mg2+, CO2, and pH', Archives of Biochemistry and Biophysics, pp. 723–729. Available at: https://doi.org/10.1016/0003-9861(76)90565-8.

Balasubramaniam, K.N. et al. (2012) 'Hierarchical steepness and phylogenetic models: Phylogenetic signals in Macaca', Animal Behaviour, 83(5), pp. 1207–1218. Available at: https://doi.org/10.1016/j.anbehav.2012.02.012.

Beerling, D.J., McElwain, J.C. and Osborne, C.P. (1998) 'Stomatal responses of the "living fossil" Ginkgo biloba L. to changes in atmospheric CO2 concentrations', Journal of Experimental Botany, 49(326), pp. 1603–1607. Available at: https://doi.org/10.1093/jxb/49.326.1603.

Beerling, D.J., Osborne, C.P. and Chaloner, W.G. (2001) 'Evolution of leaf-form in land plants linked to atmospheric CO2 decline in the Late Palaeozoic era', Nature, 410(6826), pp. 352–354. Available at: https://doi.org/10.1038/35066546.

Birky, J. (2001) 'The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models', Annual Review of Genetics, 35, pp. 125–148. Available at: https://doi.org/10.1146/annurev.genet.35.102401.090231.

Blomberg, S.P. and Garland, T. (2002) ‘Tempo and mode in evolution: Phylogenetic inertia, adaptation and comparative methods’, Journal of Evolutionary Biology, 15(6), pp. 899–910. Available at: https://doi.org/10.1046/j.1420-9101.2002.00472.x.

Blomberg, S.P., Garland, T. and Ives, A.R. (2003) ‘Testing for phylogenetic signal in comparative data: Behavioral traits are more labile’, Evolution, 57(4), pp. 717–745. Available at: https://doi.org/10.1111/j.0014-3820.2003.tb00285.x.

Blyth, C.R. (1972) ‘On Simpson’s Paradox and the Sure-Thing Principle’, Journal of the American Statistical Association, 67(338), p. 364. Available at: https://doi.org/10.2307/2284382.

Bouvier, J.W. et al. (2021) ‘Rubisco Adaptation Is More Limited by Phylogenetic Constraint Than by Catalytic Trade-off’, Molecular Biology and Evolution, 38(7), pp. 2880–2896. Available at: https://doi.org/10.1093/molbev/msab079.

Bouvier, J.W., Emms, D.M. and Kelly, S. (2024) ‘Rubisco is evolving for improved catalytic efficiency and CO2 assimilation in plants’, Proceedings of the National Academy of Sciences of the United States of America, 121(11). Available at: https://doi.org/10.1073/pnas.2321050121.

Bouvier, J.W. and Kelly, S. (2023) ‘Response to Tcherkez and Farquhar: Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off’, Journal of Plant Physiology, 287, p. 154021. Available at: https://doi.org/10.1016/j.jplph.2023.154021.

Bouvier, J.W. and Kelly, S. (2024) ‘Metabolic engineering of stomatal precursor cells enhances photosynthetic water-use efficiency and vegetative growth under water-deficit conditions in Arabidopsis thaliana’, bioRxiv [Preprint]. Available at: https://doi.org/doi.org/10.1101/2024.08.29.610053.

Brown, T.A. (2002) ‘Chapter 16 - Molecular Phylogenetics’, in Genomes, pp. 1–43.

von Caemmerer, S. et al. (1994) ‘The kinetics of ribulose-1,5-bisphosphate carboxylase/oxygenase in vivo inferred from measurements of photosynthesis in leaves of transgenic tobacco’, Planta, 195(1), pp. 88–97. Available at: https://doi.org/10.1007/BF00206296.

Caine, R.S. et al. (2019) ‘Rice with reduced stomatal density conserves water and has improved drought tolerance under future climate conditions’, New Phytologist, 221(1), pp. 371–384. Available at: https://doi.org/10.1111/nph.15344.

Capó-Bauçà, S., Iñiguez, C. and Galmés, J. (2024) ‘The diversity and coevolution of Rubisco and CO2 concentrating mechanisms in marine macrophytes’, New Phytologist, 241(6), pp. 2353–2365. Available at: https://doi.org/10.1111/nph.19528.

Carlin, D.A. et al. (2016) ‘Kinetic characterization of 100 glycoside hydrolase mutants enables the discovery of structural features correlated with kinetic constants’, PLoS ONE, 11(1). Available at: https://doi.org/10.1371/journal.pone.0147596.

Chollet, R. and Anderson, L.L. (1976) ‘Regulation of ribulose 1,5-bisphosphate carboxylase-oxygenase activities by temperature pretreatment and chloroplast metabolites’, Archives of Biochemistry and Biophysics, pp. 344–351. Available at: https://doi.org/10.1016/0003-9861(76)90173-9.

Cleton-Jansen, A.M. et al. (1991) ‘A single amino acid substitution changes the substrate specificity of quinoprotein glucose dehydrogenase in Gluconobacter oxydans’, MGG Molecular & General Genetics, 229(2), pp. 206–212. Available at: https://doi.org/10.1007/BF00272157.

Conlan, B. and Whitney, S. (2018) ‘Preparing Rubisco for a tune up’, Nature Plants, 4(1), pp. 12–13. Available at: https://doi.org/10.1038/s41477-017-0089-2.

Cummins, P.L. (2021) ‘The Coevolution of RuBisCO, Photorespiration, and Carbon Concentrating Mechanisms in Higher Plants’, Frontiers in Plant Science, 12. Available at: https://doi.org/10.3389/fpls.2021.662425.

Damborsky, J. and Brezovsky, J. (2014) ‘Computational tools for designing and engineering enzymes’, Current Opinion in Chemical Biology, pp. 8–16. Available at: https://doi.org/10.1016/j.cbpa.2013.12.003.

Doolittle, W.F. (2000) ‘Uprooting the tree of life.’, Scientific American, 282(2), pp. 90–95. Available at: https://doi.org/10.1038/scientificamerican0200-90.

Dow, G.J., Berry, J.A. and Bergmann, D.C. (2017) ‘Disruption of stomatal lineage signaling or transcriptional regulators has differential effects on mesophyll development, but maintains coordination of gas exchange’, New Phytologist, 216(1), pp. 69–75. Available at: https://doi.org/10.1111/NPH.14746.

Espadaler, J. et al. (2008) ‘Prediction of enzyme function by combining sequence similarity and protein interactions’, BMC Bioinformatics, 9. Available at: https://doi.org/10.1186/1471-2105-9-249.

Evans, J.R. et al. (2009) ‘Resistances along the CO2 diffusion pathway inside leaves’, Journal of Experimental Botany, 60(8), pp. 2235–2248. Available at: https://doi.org/10.1093/jxb/erp117.

Falquetto-Gomes, P. et al. (2024) ‘From epidermal cells to functional pores: Understanding stomatal development’, Journal of Plant Physiology, 292. Available at: https://doi.org/10.1016/j.jplph.2023.154163.

Farquhar, G.D., von Caemmerer, S. and Berry, J.A. (1980) ‘A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species’, Planta, 149(1), pp. 78–90. Available at: https://doi.org/10.1007/BF00386231.

Farquhar, G.D. and Richards, R.A. (1984) ‘Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes.’, Australian Journal of Plant Physiology, 11(6), pp. 539–552. Available at: https://doi.org/10.1071/PP9840539.

Feitosa-Araujo, E. et al. (2020) ‘Changes in intracellular NAD status affect stomatal development in an abscisic acid-dependent manner’, Plant Journal, 104(5), pp. 1149–1168. Available at: https://doi.org/10.1111/tpj.15000.

Felsenstein, J. (1985) Phylogenies and the comparative method., American Naturalist. Available at: https://doi.org/10.1086/284325.

Flamholz, A.I. et al. (2019) ‘Revisiting Trade-offs between Rubisco Kinetic Parameters’, Biochemistry, pp. 3365–3376. Available at: https://doi.org/10.1021/acs.biochem.9b00237.

Flexas, J. et al. (2013) ‘Diffusional conductances to CO2 as a target for increasing photosynthesis and photosynthetic water-use efficiency’, Photosynthesis Research, 117(1–3), pp. 45–59. Available at: https://doi.org/10.1007/s11120-013-9844-z.

Franks, P.J. and Beerling, D.J. (2009) ‘Maximum leaf conductance driven by CO2 effects on stomatal size and density over geologic time’, Proceedings of the National Academy of Sciences of the United States of America, 106(25), pp. 10343–10347. Available at: https://doi.org/10.1073/pnas.0904209106.

Freckleton, R.P., Harvey, P.H. and Pagel, M. (2002) ‘Phylogenetic analysis and comparative data: A test and review of evidence’, American Naturalist, 160(6), pp. 712–726. Available at: https://doi.org/10.1086/343873.

Fritz, S.A. and Purvis, A. (2010) ‘Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits’, Conservation Biology, 24(4), pp. 1042–1051. Available at: https://doi.org/10.1111/j.1523-1739.2010.01455.x.

Galmés, J. et al. (2014) ‘Expanding knowledge of the Rubisco kinetics variability in plant species: Environmental and evolutionary trends’, Plant, Cell and Environment, 37(9), pp. 1989–2001. Available at: https://doi.org/10.1111/pce.12335.

Galmés, J. et al. (2015) ‘Temperature responses of the Rubisco maximum carboxylase activity across domains of life: Phylogenetic signals, trade-offs, and importance for carbon gain’, Photosynthesis Research, 123(2), pp. 183–201. Available at: https://doi.org/10.1007/s11120-014-0067-8.

Gielly, L. and Taberlet, P. (1994) The use of chloroplast DNA to resolve plant phylogenies: Noncoding versus rbcL sequences, Molecular Biology and Evolution. Available at: https://doi.org/10.1093/oxfordjournals.molbev.a040157.

Gionfriddo, M., Rhodes, T. and Whitney, S.M. (2023) ‘Perspectives on improving crop Rubisco by directed evolution’, Seminars in Cell and Developmental Biology [Preprint]. Available at: https://doi.org/10.1016/j.semcdb.2023.04.003.

Gionfriddo, M., Zang, K. and Hayer-Hartl, M. (2023) ‘The challenge of engineering Rubisco for improving photosynthesis’, FEBS Letters, 597(13), pp. 1679–1680. Available at: https://doi.org/10.1002/1873-3468.14678.

Gittleman, J.L. and Kot, M. (1990) ‘Adaptation: Statistics and a Null Model for Estimating Phylogenetic Effects’, Systematic Biology, 39(3), pp. 227–241. Available at: https://doi.org/10.2307/2992183.

Goudet, M.M.M. et al. (2020) ‘Rubisco and carbon-concentrating mechanism co-evolution across chlorophyte and streptophyte green algae’, New Phytologist, 227(3), pp. 810–823. Available at: https://doi.org/10.1111/nph.16577.

Gutteridge, S. et al. (1984) ‘ A site-specific mutation within the active site of ribulose-1,5-bisphosphate carboxylase of Rhodospirillum rubrum ’, The EMBO Journal, 3(12), pp. 2737–2743. Available at: https://doi.org/10.1002/j.1460-2075.1984.tb02204.x.

Gutteridge, S., Lorimer, G. and Pierce, J. (1988) ‘Details of the reactions catalysed by mutant forms of rubisco’, Plant Physiol Biochem (Paris), 26, pp. 675–682.

Hansen, T.F. and Martins, E.P. (1996) ‘Translating between microevolutionary process and macroevolutionary patterns: The correlation structure of interspecific data’, Evolution, 50(4), pp. 1404–1417. Available at: https://doi.org/10.1111/j.1558-5646.1996.tb03914.x.

Harrison, E.L. et al. (2020) ‘The influence of stomatal morphology and distribution on photosynthetic gas exchange’, Plant Journal, 101(4), pp. 768–779. Available at: https://doi.org/10.1111/tpj.14560.

Hermida-Carrera, C., Kapralov, M. V. and Galmés, J. (2016) ‘Rubisco catalytic properties and temperature response in crops’, Plant Physiology, 171(4), pp. 2549–2561. Available at: https://doi.org/10.1104/pp.16.01846.

Hughes, J. et al. (2017) ‘Reducing stomatal density in barley improves drought tolerance without impacting on yield’, Plant Physiology, 174(2), pp. 776–787. Available at: https://doi.org/10.1104/pp.16.01844.

Igamberdiev, A.U. and Gardeström, P. (2003) ‘Regulation of NAD- and NADP-dependent isocitrate dehydrogenases by reduction levels of pyridine nucleotides in mitochondria and cytosol of pea

leaves', Biochimica et Biophysica Acta - Bioenergetics, 1606(1–3), pp. 117–125. Available at: https://doi.org/10.1016/S0005-2728(03)00106-3.

Iñiguez, C. et al. (2020) 'Evolutionary trends in RuBisCO kinetics and their co-evolution with CO2 concentrating mechanisms', Plant Journal, 101(4), pp. 897–918. Available at: https://doi.org/10.1111/tpj.14643.

Iqbal, W.A., Lisitsa, A. and Kapralov, M. V (2022) 'Predicting plant Rubisco kinetics from RbcL sequence data using machine learning', Journal of Experimental Botany [Preprint]. Available at: https://doi.org/10.1093/jxb/erac368.

Ives, A.R. (2019) 'R 2 s for Correlated Data: Phylogenetic Models, LMMs, and GLMMs', Systematic Biology, 68(2), pp. 234–251. Available at: https://doi.org/10.1093/sysbio/syy060.

James, S.A. and Bell, D.T. (2000) 'Influence of light availability on leaf structure and growth of two Eucalyptus globulus ssp. globulus provenances', Tree Physiology, 20(15), pp. 1007–1018. Available at: https://doi.org/10.1093/treephys/20.15.1007.

Jenkins, C.L., Furbank, R.T. and Hatch, M.D. (1989) 'Mechanism of c(4) photosynthesis: a model describing the inorganic carbon pool in bundle sheath cells.', Plant physiology, 91(4), pp. 1372–81. Available at: https://doi.org/10.1104/pp.91.4.1372.

Johnson, E.T. et al. (2001) 'Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase', Plant Journal, 25(3), pp. 325–333. Available at: https://doi.org/10.1046/j.1365-313X.2001.00962.x.

Jordan, D.B. and Ogren, W.L. (1981) 'Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase', Nature, 291(5815), pp. 513–515. Available at: https://doi.org/10.1038/291513a0.

Jordan, D.B. and Ogren, W.L. (1983) 'Species variation in kinetic properties of ribulose 1,5-bisphosphate carboxylase/oxygenase', Archives of Biochemistry and Biophysics, 227(2), pp. 425–433. Available at: https://doi.org/10.1016/0003-9861(83)90472-1.

Jumper, J. et al. (2021) 'Highly accurate protein structure prediction with AlphaFold', Nature, 596(7873), pp. 583–589. Available at: https://doi.org/10.1038/s41586-021-03819-2.

Kamilar, J.M. and Cooper, N. (2013) 'Phylogenetic signal in primate behaviour, ecology and life history', Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1618). Available at: https://doi.org/10.1098/rstb.2012.0341.

Kapralov, M. V. and Filatov, D.A. (2007) 'Widespread positive selection in the photosynthetic Rubisco enzyme', BMC Evolutionary Biology, 7. Available at: https://doi.org/10.1186/1471-2148-7-73.

Kellogg, E.A. and Juliano, N.D. (1997) 'The structure and function of RuBisCO and their implications for systematic studies', American Journal of Botany, 84(3), pp. 413–428. Available at: https://doi.org/10.2307/2446015.

Kelly, S., Grenyer, R. and Scotland, R.W. (2014) 'Phylogenetic trees do not reliably predict feature diversity', Diversity and Distributions, 20(5), pp. 600–612. Available at: https://doi.org/10.1111/ddi.12188.

Klepikova, A. V. et al. (2016) 'A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling', Plant Journal, 88(6), pp. 1058–1070. Available at: https://doi.org/10.1111/tpj.13312.

Kubien, D.S. et al. (2008) ‘The biochemistry of Rubisco in Flaveria’, Journal of Experimental Botany, 59(7), pp. 1767–1777. Available at: https://doi.org/10.1093/jxb/erm283.

Lake, J.A. et al. (2001) ‘Plant development. Signals from mature to new leaves.’, Nature, 411(6834).

Lim, S.L. et al. (2020) ‘In planta study of photosynthesis and photorespiration using NADPH and NADH/NAD+ fluorescent protein sensors’, Nature Communications, 11(1). Available at: https://doi.org/10.1038/s41467-020-17056-0.

Lin, M.T. et al. (2022) ‘Improving the efficiency of Rubisco by resurrecting its ancestors in the family Solanaceae’, Science Advances, 8(15). Available at: https://doi.org/10.1126/sciadv.abm6871.

Lorimer, G.H. and Andrews, T.J. (1973) ‘Plant photorespiration - An inevitable consequence of the existence of atmospheric oxygen’, Nature, 243(5406), pp. 359–360. Available at: https://doi.org/10.1038/243359a0.

Losos, J.B. (2008) ‘Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species’, Ecology Letters, pp. 995–1003. Available at: https://doi.org/10.1111/j.1461-0248.2008.01229.x.

Lukas, D., Towner, M. and Borgerhoff Mulder, M. (2021) ‘The potential to infer the historical pattern of cultural macroevolution’, Philosophical Transactions of the Royal Society B: Biological Sciences, 376(1828). Available at: https://doi.org/10.1098/rstb.2020.0057.

Lundgren, M.R. et al. (2019) ‘Mesophyll porosity is modulated by the presence of functional stomata’, Nature Communications, 10(1). Available at: https://doi.org/10.1038/s41467-019-10826-5.

McElwain, J.C. and Chaloner, W.G. (1995) ‘Stomatal density and index of fossil plants track atmospheric carbon dioxide in the palaeozoic’, Annals of Botany, 76(4), pp. 389–395. Available at: https://doi.org/10.1006/anbo.1995.1112.

Meyer, M. and Griffiths, H. (2013) ‘Origins and diversity of eukaryotic CO2-concentrating mechanisms: Lessons for the future’, Journal of Experimental Botany, pp. 769–786. Available at: https://doi.org/10.1093/jxb/ers390.

Minshull, J. et al. (2005) ‘Predicting enzyme function from protein sequence’, Current Opinion in Chemical Biology, pp. 202–209. Available at: https://doi.org/10.1016/j.cbpa.2005.02.003.

Mitra, A., Choi, H.K. and An, G. (1989) ‘Structural and functional analyses of Arabidopsis thaliana chlorophyll a/b-binding protein (cab) promoters’, Plant Molecular Biology, 12(2), pp. 169–179. Available at: https://doi.org/10.1007/BF00020502.

Movahedi, M. et al. (2021) ‘Stomatal responses to carbon dioxide and light require abscisic acid catabolism in Arabidopsis’, Interface Focus, 11(2). Available at: https://doi.org/10.1098/rsfs.2020.0036.

Nutman, A.P. et al. (2016) ‘Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures’, Nature, 537(7621), pp. 535–538. Available at: https://doi.org/10.1038/nature19355.

Occhialini, A. et al. (2016) ‘Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO2’, Plant Journal, 85(1), pp. 148–160. Available at: https://doi.org/10.1111/tpj.13098.

Orr, D.J. et al. (2016) ‘Surveying rubisco diversity and temperature response to improve crop photosynthetic efficiency’, Plant Physiology, 172(2), pp. 707–717. Available at: https://doi.org/10.1104/pp.16.00750.

Pagel, M. (1999) ‘Inferring the historical patterns of biological evolution’, Nature. Nature Publishing Group, pp. 877–884. Available at: https://doi.org/10.1038/44766.

Parry, M.A.J., Keys, A.J. and Gutteridge, S. (1989) ‘Variation in the specificity factor of C3 higher plant rubiscos determined by the total consumption of ribulose-P2’, Journal of Experimental Botany, 40(3), pp. 317–320. Available at: https://doi.org/10.1093/jxb/40.3.317.

Prins, A. et al. (2016) ‘Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis’, Journal of Experimental Botany, 67(6), pp. 1827–1838. Available at: https://doi.org/10.1093/jxb/erv574.

Prywes, N. et al. (2022) ‘Rubisco function, evolution, and engineering’. Available at: http://arxiv.org/abs/2207.10773.

Robbins, E.H.J. and Kelly, S. (2023) ‘The Evolutionary Constraints on Angiosperm Chloroplast Adaptation’, Genome Biology and Evolution, 15(6). Available at: https://doi.org/10.1093/gbe/evad101.

Sage, R.F. (2002) ‘Variation in the kcat of Rubisco in C3 and C4 plants and some implications for photosynthetic performance at high and low temperature’, Journal of Experimental Botany, pp. 609–620. Available at: https://doi.org/10.1093/jexbot/53.369.609.

Sally Buck et al. (2023) ‘Escherichia coli expressing chloroplast chaperones as a proxy to test heterologous Rubisco production in leaves’, Journal of Experimental Botany, pp. 664–676.

Savir, Y. et al. (2010) ‘Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape’, Proceedings of the National Academy of Sciences of the United States of America, 107(8), pp. 3475–3480. Available at: https://doi.org/10.1073/pnas.0911663107.

Schomburg, I. et al. (2017) ‘The BRENDA enzyme information system–From a database to an expert system’, Journal of Biotechnology, 261, pp. 194–206. Available at: https://doi.org/10.1016/j.jbiotec.2017.04.020.

Schomburg, I., Chang, A. and Schomburg, D. (2002) ‘BRENDA, enzyme data and metabolic information’, Nucleic Acids Research, 30(1), pp. 47–49. Available at: https://doi.org/10.1093/nar/30.1.47.

Seemann, J.R., Badger, M.R. and Berry, J.A. (1984) ‘Variations in the Specific Activity of Ribulose-1,5-bisphosphate Carboxylase between Species Utilizing Differing Photosynthetic Pathways’, Plant Physiology, 74(4), pp. 791–794. Available at: https://doi.org/10.1104/pp.74.4.791.

Sharwood, R.E. et al. (2016) ‘Temperature responses of Rubisco from Paniceae grasses provide opportunities for improving C 3 photosynthesis’, Nature Plants, 2. Available at: https://doi.org/10.1038/nplants.2016.186.

Sharwood, R.E. (2017) ‘Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops’, New Phytologist, pp. 494–510. Available at: https://doi.org/10.1111/nph.14351.

Siegel, J.B. et al. (2015) ‘Computational protein design enables a novel one-carbon assimilation pathway’, Proceedings of the National Academy of Sciences of the United States of America, 112(12), pp. 3704–3709. Available at: https://doi.org/10.1073/pnas.1500545112.

Simpson, G.G. (1959) ‘Anatomy and morphology: classification and evolution: 1859 and 1959’, Proceedings of the American Philosophical Society, (2), pp. 286–306. Available at: http://www.jstor.org/stable/10.2307/985154.

Smith, D.R. (2015) ‘Mutation rates in plastid genomes: They are lower than you might think’, Genome Biology and Evolution, 7(5), pp. 1227–1234. Available at: https://doi.org/10.1093/gbe/evv069.

de Souza Chaves, I. et al. (2019) ‘The mitochondrial NAD+ transporter (NDT1) plays important roles in cellular NAD+ homeostasis in Arabidopsis thaliana’, Plant Journal, 100(3), pp. 487–504. Available at: https://doi.org/10.1111/tpj.14452.

Tcherkez, G. and Farquhar, G.D. (2021) ‘Rubisco catalytic adaptation is mostly driven by photosynthetic conditions – Not by phylogenetic constraints’, Journal of Plant Physiology, 267. Available at: https://doi.org/10.1016/j.jplph.2021.153554.

Tcherkez, G.G.B., Farquhar, G.D. and Andrews, T.J. (2006) ‘Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized’, Proceedings of the National Academy of Sciences of the United States of America, 103(19), pp. 7246–7251. Available at: https://doi.org/10.1073/pnas.0600605103.

Tessler, M. et al. (2022) ‘Let’s end taxonomic blank slates with molecular morphology’, Frontiers in Ecology and Evolution, 10. Available at: https://doi.org/10.3389/fevo.2022.1016412.

Tichá, I. (1982) ‘Photosynthetic characteristics during ontogenesis of leaves: 7. Stomata density and sizes’, Photosynthetica, 16(2), p. 375—471.

Vigna, L. et al. (1980) ‘Dependence of the Stomatal Index on Environmental Factors during Stomatal Differentiation in Leaves of Vigna sinensis L.’, Journal of Experimental Botany, 31(124), pp. 1211–1216.

Villar, K. Del et al. (1997) ‘Amino acid substitutions that convert the protein substrate specificity of farnesyltransferase to that of geranylgeranyltransferase type I’, Journal of Biological Chemistry, 272(1), pp. 680–687. Available at: https://doi.org/10.1074/jbc.272.1.680.

Wessinger, M.E., Edwards, G.E. and Ku, M.S.B. (1989) ‘Quantity and kinetic properties of ribulose 1,5-bisphosphate carboxylase in C3, C4, and C3-C4 intermediate species of flaveria (asteraceae)’, Plant and Cell Physiology, 30(5), pp. 665–671.

Wilson, R.H. et al. (2018) ‘An improved Escherichia coli screen for Rubisco identifies a protein-protein interface that can enhance CO2-fixation kinetics’, Journal of Biological Chemistry, 293(1), pp. 18–27. Available at: https://doi.org/10.1074/jbc.M117.810861.

Wilson, R.H., Alonso, H. and Whitney, S.M. (2016) ‘Evolving Methanococcoides burtonii archaeal Rubisco for improved photosynthesis and plant growth’, Scientific Reports, 6. Available at: https://doi.org/10.1038/srep22284.

Wolfe, K.H., Li, W.H. and Sharp, P.M. (1987) ‘Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs.’, Proceedings of the National Academy of Sciences of the United States of America, 84(24), pp. 9054–9058. Available at: https://doi.org/10.1073/pnas.84.24.9054.

Woodward, F.I. (1987) ‘Stomatal numbers are sensitive to increases in CO2 from pre-industrial levels’, Nature, 327, pp. 617–618.

Woodward, F.I. and Bazzaz, F.A. (1988) ‘The responses of stomatal density to CO2 partial pressure’, Journal of Experimental Botany, 39(12), pp. 1771–1781. Available at: https://doi.org/10.1093/jxb/39.12.1771.

Woodward, F.I., Lake, J.A. and Quick, W.P. (2002) ‘Stomatal development and CO2: Ecological consequences’, New Phytologist, 153(3), pp. 477–484. Available at: https://doi.org/10.1046/j.0028-646X.2001.00338.x.

Wu, E.T.Y. et al. (2021) ‘Detecting the phylogenetic signal of glacial refugia in a bryodiversity hotspot outside the tropics’, Diversity and Distributions [Preprint]. Available at: https://doi.org/10.1111/ddi.13449.

Yang, Z. and Rannala, B. (2012) ‘Molecular phylogenetics: Principles and practice’, Nature Reviews Genetics, 13(5), pp. 303–314. Available at: https://doi.org/10.1038/nrg3186.

Yeoh, H.-H., Badger, M.R. and Watson, L. (1980) ‘ Variations in K m (CO 2) of Ribulose-1,5-bisphosphate Carboxylase among Grasses ’, Plant Physiology, 66(6), pp. 1110–1112. Available at: https://doi.org/10.1104/pp.66.6.1110.

Yeoh, H.-H., Badger, M.R. and Watson, L. (1981) ‘Variations in Kinetic Properties of Ribulose-1,5-bisphosphate Carboxylases among Plants’, Plant Physiology, 67(6), pp. 1151–1155. Available at: https://doi.org/10.1104/pp.67.6.1151.

Young, J.N. et al. (2016) ‘Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentrating mechanisms’, Journal of Experimental Botany, 67(11), pp. 3445–3456. Available at: https://doi.org/10.1093/jxb/erw163.

Young, J.N. and Hopkinson, B.M. (2017) ‘The potential for co-evolution of CO2-concentrating mechanisms and Rubisco in diatoms’, Journal of Experimental Botany, 68(14), pp. 3751–3762. Available at: https://doi.org/10.1093/jxb/erx130.

Zhou, Y. and Whitney, S. (2019) ‘Directed evolution of an improved Rubisco; in vitro analyses to decipher fact from fiction’, International Journal of Molecular Sciences, 20(20). Available at: https://doi.org/10.3390/ijms20205019.

Zhu, G. et al. (1998) ‘Dependence of catalysis and CO2/O2 specificity of Rubisco on the carboxy-terminus of the large subunit at different temperatures’, Photosynthesis Research, 57(1), pp. 71–79. Available at: https://doi.org/10.1023/A:1006020424581.

Appendix A, Supplemental Chapter: Response to Tcherkez and Farquhar: Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off

Appendix A, Contributors:
Bouvier, J.W. and Kelly, S.

Appendix A, Author contributions:
J.W.B. and S.K. conceived the study, performed the computational analysis, and wrote the manuscript.

This thesis chapter has been published at the following location:
Bouvier, J.W. and Kelly, S. 2023. Journal of Plant Physiology, Article No 154021

All supplemental data files are provided in the online version of this manuscript.

A.1 Abstract

Rubisco is the primary entry point for carbon into the biosphere. It has been widely proposed that rubisco is highly constrained by catalytic trade-offs due to correlations between the enzyme's kinetic traits across species. In previous work, we have shown that these correlations, and thus the strength of catalytic trade-offs, have been overestimated due to the presence of phylogenetic signal in the kinetic trait data (Bouvier et al., 2021). We demonstrated that only the trade-offs between the Michaelis constant for CO_2 and carboxylase turnover, and between the Michaelis constants for CO_2

and O_2 were robust to phylogenetic effects. We further demonstrated that phylogenetic constraints have limited rubisco adaptation to a greater extent than the combined action of catalytic trade-offs. Recently, however, our claims have been contested by Tcherkez and Farquhar (2021), who have argued that the phylogenetic signal we detect in rubisco kinetic traits is an artefact of species sampling, the use of *rbcL*-based trees for phylogenetic inference, laboratory-to-laboratory variability in kinetic measurements, and homoplasy of the C_4 trait. In the present article, we respond to these criticisms on a point-by-point basis and conclusively show that all are unfounded. As such, we stand by our original conclusions. Namely, that rubisco kinetic evolution has been limited by biochemical trade-offs, but that these are not absolute and have been previously overestimated due to phylogenetic biases. Instead, rubisco adaptation has in fact been more limited by phylogenetic constraint.

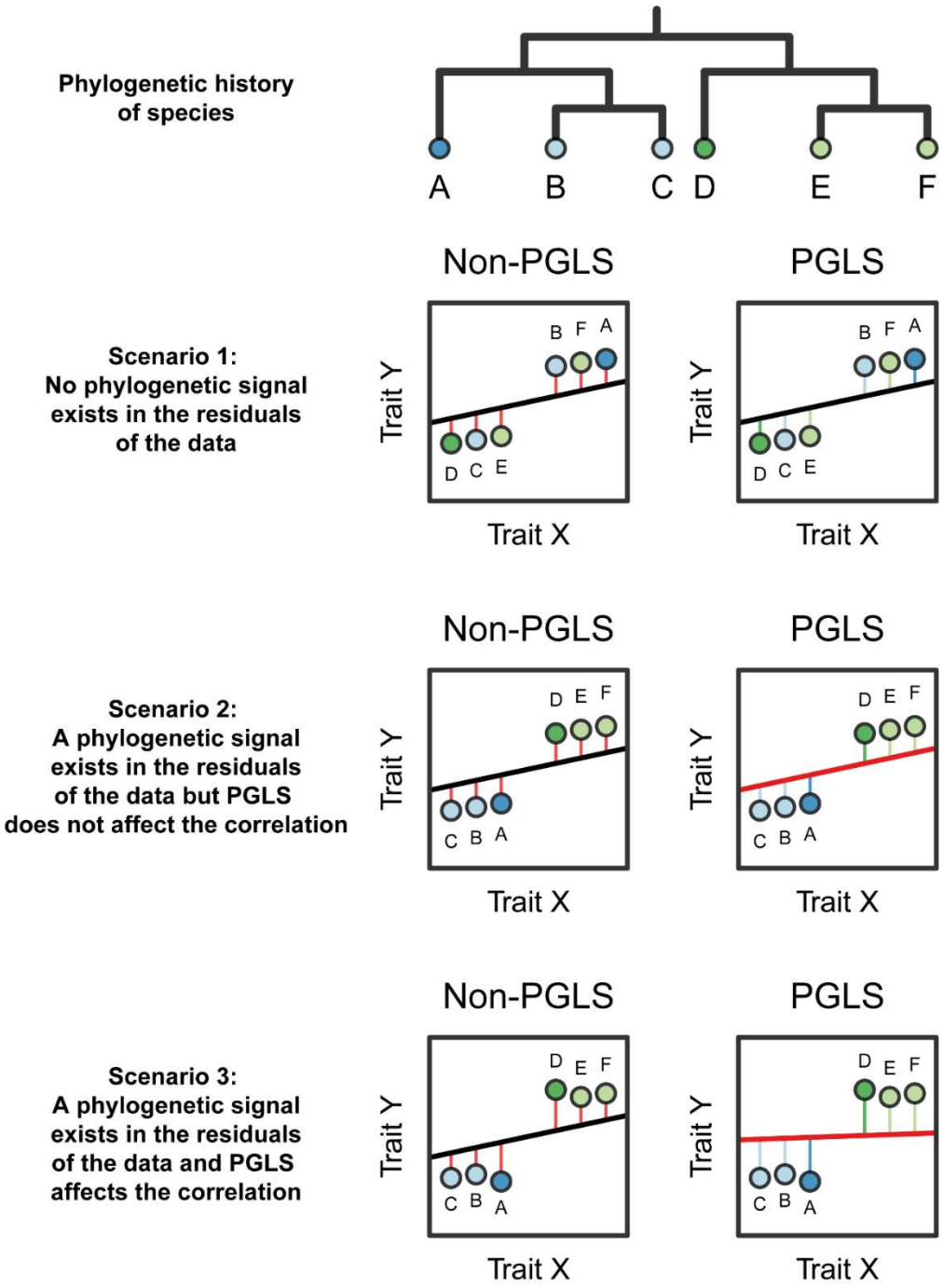
A.2 Introduction

Phylogenetic signal is the phenomenon by which closely related individuals resemble each other more than individuals that are distantly related (Blomberg et al., 2003; Blomberg and Garland, 2002). In biological data, the presence of phylogenetic signal is determined by evaluating the covariation in traits across species in the context of their proximity to one another on a phylogenetic tree. Although such a phylogenetic structure in interspecific trait variation might be expected *a priori* based on the action of inheritance and descent with modification, phylogenetic signal is not universal in traits or across taxonomic groups (Blomberg et al., 2003; Freckleton et al., 2002; Kamilar and Cooper, 2013; Losos, 2008). For instance, many evolutionary processes including rapid adaptation, divergent selection, and convergent evolution can lead to significant departures from phylogenetic signal by either causing similarity among species to be independent of evolutionary distance, or by causing distantly related species to appear more similar on average compared to those which share more recent common ancestry (Hansen and Martins, 1996; Kamilar and Cooper, 2013; Kelly, Grenyer and Scotland, 2014). Moreover, although genetic sequence similarity is used to infer evolutionary history among species, it can be difficult to predict the molecular basis of trait variation. In enzymes, for example, single point mutations can have dramatic consequences for catalytic function (Cleton-

Jansen et al., 1991; Johnson et al., 2001; Villar et al., 1997) whilst divergent sequences can maintain similar biochemical properties (Espadaler et al., 2008). As such, studying phylogenetic signal can be instructive to interpret patterns of biological diversity and their underpinning evolutionary processes, as well as to help map the sequence–function landscape to better understand genotype–phenotype interactions.

Aside from the fundamental interest in examining the phylogenetic basis of trait data, quantification of phylogenetic signal is also of importance in comparative biology. This is because the shared ancestry of all taxa as determined by the hierarchical tree of life causes a "serious statistical problem" (Felsenstein, 1985) in cross-species analyses. Specifically, this problem arises due to the fact that biological datapoints are not independent observations, as is a core assumption of conventional statistical methods (Appendix A, table S1). Instead, the phylogenetic non-independence between taxa means that the trait variation captured by a given dataset is an artefact of species sampling, where the magnitude of this introduced sampling bias is affected by the strength and pattern of phylogenetic signal in the traits of interest. Nevertheless, although this non-independence of species violates the assumptions of conventional statistical methods, specialist phylogenetically aware methods have been developed to circumvent this issue (Freckleton et al., 2002; Lindenfors et al., 2010; Motani and Schmitz, 2011; Revell, 2009). In contrast to their non-phylogenetic counterparts, these phylogenetic methods do not assume independence among data points, but instead measure and account for the relative non-independence of all species in the context of their phylogenetic relatedness (Appendix A, table S1). In this way, the invention of phylogenetic statistical tools has enabled biologically meaningful conclusions to be drawn from cross-species analyses.

One of the most common phylogenetic statistical methods, the phylogenetic generalized least squares regression (Grafen, 1989; Martins and Hansen, 1997; Pagel, 1999, 1997; Rohlf, 2001), assesses the association between continuous biological traits using an analogous approach to a conventional least squares regression. However, in contrast to a conventional regression which assumes independence and thus weights all datapoints equally, the phylogenetic regression works by weighting the importance of data differentially depending on the degree of shared ancestry of



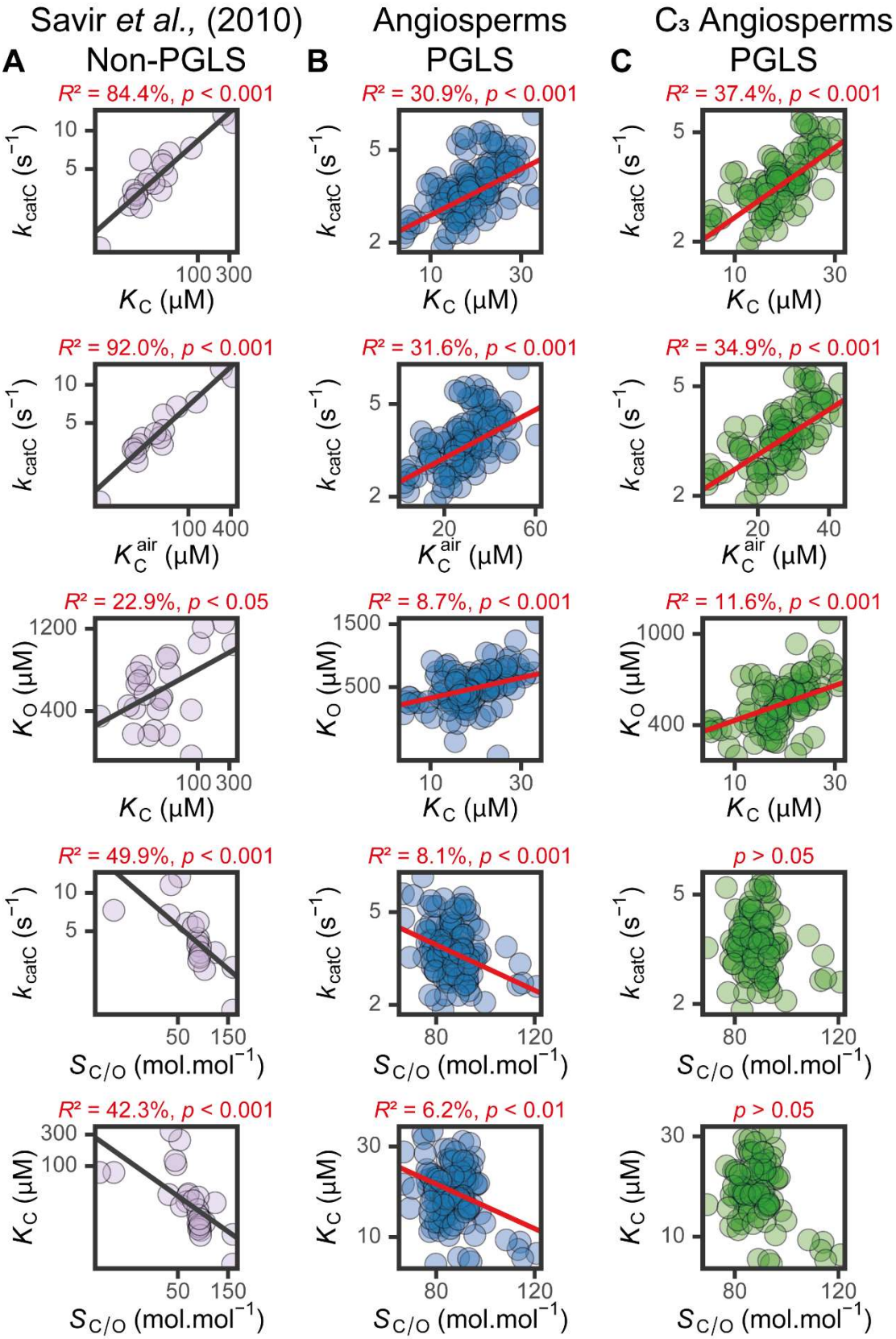
Appendix A, Figure S1. Cartoon to illustrate how conventional and phylogenetic generalized least squares regression models infer relationships between biological traits. Under all scenarios 1 – 3, conventional non-phylogenetic regression models are blind to the underlying phylogenetic history of the sampled species and compute the relationship between trait X and trait Y assuming all datapoints are independent observations. In contrast, phylogenetic regression models measure the co-variation between species relatedness and their position in X~Y trait space, subsequently taking this information into account during downstream model fitting. Specifically, if there is no phylogenetic signal in the residuals of the data, a phylogenetic correction of the data is not necessary and a conventional regression model will be fitted (Scenario 1). Alternately, if a phylogenetic signal in the residual errors is detected, a phylogenetic correction of the data will be performed where depending

on the data, this may or may not affect the observed relationship relative to if a non-phylogenetic regression model was fitted (Scenario 2 and 3). These scenarios are illustrated using trait data from six species (identified by letters A—F). The evolutionary relationships among these species are described based on the phylogenetic tree and are also visualized by the colours of each species datapoint. Lines of best fit from conventional and phylogenetic generalized least squares regression models are shown in black and red, respectively. **[*End of Figure Legend*].**

---

species with one another (Appendix A, Figure S1). For an excellent review of the conceptual and technical background to this tool, see (Symonds and Blomberg, 2014). In brief, this feat is achieved by incorporating a phylogenetic parameter in the model formulation which accounts for the phylogenetic signal in the residual errors of the data (i.e., the co-variation between species relatedness and their position in X~Y trait space). If no phylogenetic signal exists in the residual errors, the phylogenetic parameter of the model is equal to 0 and a standard non-phylogenetic regression is performed (Appendix A, Figure S1). In contrast, as the phylogenetic signal in the residual errors increases, this is accordingly controlled for in the model by reducing the weight of data from groups of closely related species on the phylogenetic tree as compared to species which are more evolutionarily distant (Appendix A, Figure S1). Therefore, in this way, phylogenetic least squares regression accounts for the non-independence of biological data and generates a robust and accurate estimate of the gradient, intercept, coefficient of variation (R^2) and associated p-values for biological associations (Appendix A, Figure S1).

The necessity to consider phylogenetic context when analyzing interspecies data has been widely adopted in many fields of biology, including ecology (e.g., Wu et al., 2021), animal behavior (e.g., Balasubramaniam et al., 2012), anthropology (e.g., Lukas, Towner and Borgerhoff Mulder, 2021) and conservation science (e.g., Fritz and Purvis, 2010). In contrast, in other sub-disciplines of biology, especially those which study phenomena at the cell or molecular level, phylogenetic statistical practices are less commonplace. Thus, it is likely that many studies which have compared quantitative data in cellular or molecular datasets, particularly those which do so between species, have been affected by the presence of phylogenetic signal.



Appendix A, Figure S2. Canonical kinetic trait trade-offs in rubisco as estimated using conventional and phylogenetic generalized least squares regression models. **A)** The pairwise correlations between rubisco kinetic traits across diverse photosynthetic organisms as proposed in the seminal cross-species analysis of (Savir et al., 2010), measured using non-phylogenetic generalized least squares regression and a small dataset of species. **B)** The pairwise correlations between rubisco

kinetic traits in angiosperms as described in (Bouvier et al., 2021), measured using phylogenetic generalized least squares regression. **C)** As in (B) but including only C_3 angiosperms. Correlation coefficients (R^2) and associated P-values between rubisco kinetic trait correlations are displayed above each respective plot. Significance values are represented as α levels, where; α = 0.001 if $P < 0.001$, α = 0.01 if $0.001 < P < 0.01$, α = 0.05 if $0.01 < P < 0.05$, and α = ns if $P > 0.05$. Both x and y axes for all plots are on a log transformed scale and respective units are shown in axis labels. Lines of best fit from conventional and phylogenetic generalized least squares regression models are shown in black and red, respectively. **[*End of Figure Legend*].**

In a recent study, we performed a cross-species analysis of the kinetic traits of the enzyme rubisco (ribulose-1,5-bisphosphate [RuBP] carboxylase/oxygenase) (Bouvier et al., 2021). Specifically, we set out to examine the constraints which have limited the enzyme's adaptation whilst correctly accounting for the phylogenetic signal arising from the non-independence of the kinetic measurements. Rubisco presents an interesting subject as the basis for such investigation as despite being the principal carbon-fixing enzyme in the biosphere (Field et al., 1998; Tabita et al., 2008), it is widely considered to be poorly optimised because it exhibits a modest rate of CO_2 turnover (Badger et al., 1998; Bar-Even et al., 2011) and catalyses a mostly counterproductive reaction with O_2 (Bowes et al., 1971; Chollet, 1977; Sharkey, 2020). The initial hypothesis which was put forward to explain this paradox of why this enzyme of paramount biological importance appears poorly adapted for its role in primary CO_2 fixation was pioneered by two comparative studies which reported severe antagonistic correlations between rubisco kinetic traits across species and proposed that these were caused by chemical constraints on the catalytic mechanism of the enzyme (Savir et al., 2010; Tcherkez et al., 2006). However, these studies, as well as all other subsequent analyses which have investigated rubisco kinetic trait correlations (Flamholz et al., 2019; Iñiguez et al., 2020; Young et al., 2016), did not account for the phylogenetic context of the species being analysed and the fact that all existing rubisco are related by evolution from a common ancestor. As such, it is possible that these correlations suffered from Felsenstein's "serious statistical problem" (Felsenstein, 1985), either in whole or in part, meaning that they are potentially an artefact of the presence of phylogenetic signal in rubisco kinetics and the non-independence of species on the phylogenetic tree.

In Bouvier et al., (2021), we found that all rubisco kinetic traits exhibit strong phylogenetic signal (Appendix A, table S2). This signal was observed when analysing the kinetic data across the tree of life, including among C_3 angiosperms, among all angiosperms, and among all photosynthetic organisms for which kinetic data were available, respectively (Appendix A, table S2). Given this observed non-independence of the kinetic data, we re-evaluated the kinetic trait correlations between species using phylogenetic least squares regression and found that all were attenuated compared to previously published values (Appendix A, Figure S2A – S2C) (Bouvier et al., 2021). Thus, phylogenetic non-independence had caused catalytic trade-offs to be overestimated in previous analyses. Despite this, we showed that there was still a genuine antagonism between the Michaelis constant for CO_2 (K_C) and carboxylase turnover (k_{catC}) (variance explained = 21–37%), and between the Michaelis constants for CO_2 and O_2 (K_O) (variance explained 9–19%) which was robust to phylogenetic correction, with all other catalytic trade-offs being either negligible or non-significant (Appendix A, Figure S2B and S2C) (Bouvier et al., 2021). Following this, we demonstrated that phylogenetic constraints explained more variation in rubisco kinetics (variance explained 30–61%), and have thus had a larger impact on limiting enzyme adaptation, compared to the combined action of all catalytic trade-offs (variance explained 6–9%) (Bouvier et al., 2021). Thus, although some rubisco catalytic trade-offs exist, these are weaker than previously thought and only partially limit rubisco adaptation in conjunction with phylogenetic constraints.

Although the above work sheds light on our understanding of the constraints that have shaped rubisco adaptation, the validity of the results presented in Bouvier et al., (2021) have been brought into question in a recent opinion piece (Tcherkez and Farquhar, 2021). In the present study, we address and refute the criticisms made by Tcherkez and Farquhar. In doing so, we confirm that the data we previously presented is valid and robust and thus our original conclusions are unaltered. Specifically, although some catalytic trade-offs are present in rubisco as determined by chemistry, the extent of this constraint on enzyme adaptation is overestimated unless phylogenetically appropriate methods are used. Instead, phylogenetic constraints have provided a more severe limitation on rubisco optimisation.

A.3 Results

A.3.1 Kinetic and phylogenetic data

As the basis of the analysis in the present study, the same rubisco kinetic dataset was used as described in Bouvier et al., (2021). This includes experimental measurements of CO_2/O_2 specificity ($S_{C/O}$), maximum carboxylase turnover rate per active site (k_{catC}), the respective Michaelis constant (i.e., the substrate concentration at half-saturated catalysed rate) for both CO_2 (K_C) and O_2 (K_O) substrates, as well as an inferred measurement of the Michaelis constant for CO_2 in 20.95% O_2 air (K_C^{air}) calculated from the formula $K_C^{air} = K_C + (K_C\,[O_2]/K_O)$, where 20.95% $[O_2]$ in water is 253 µM. As before, statistical interrogation of this dataset was restricted to log transformed trait values to ensure all data conformed to the distribution assumptions of the statistical analysis herein. Moreover, consistent with the analysis in Bouvier et al., (2021), the same phylogenetic trees of studied species were used (unless explicitly stated) as previously generated from the coding sequences of the rubisco large subunit (*rbcL*) gene.

To assess whether rubisco in different angiosperms display similar kinetics as a consequence of their phylogenetic relationship, the presence and magnitude of phylogenetic signal in rubisco kinetic traits was assessed using the statistical tools previously applied in Bouvier et al., (2021), including Pagel's lambda (Pagel, 1999), Blomberg's *K* (Blomberg et al., 2003), Blomberg's *K** (Blomberg et al., 2003), Moran's *I* (Gittleman and Kot, 1990) and Abouheif's *Cmean* (Abouheif, 1999). For an overview of the inherent differences between these methods utilized see (Bouvier et al., 2021), or for more extensive discussion see (Münkemüller et al., 2012). However, in brief, signal strength for each metric typically varies between 0 (absence of phylogenetic signal) and 1 (strong phylogenetic signal), and we assess the presence or absence of a phylogenetic signal in each trait as the majority result across these methods (i.e., the consensus result in ≥3 out of 5 methods tested).

A.3.2 A phylogenetic signal cannot be generated with a randomly simulated arbitrary trait

The first criticism of our work made by Tcherkez and Farquhar (2021) states that a phylogenetic signal may be generated using an arbitrary trait that is randomly distributed across species.

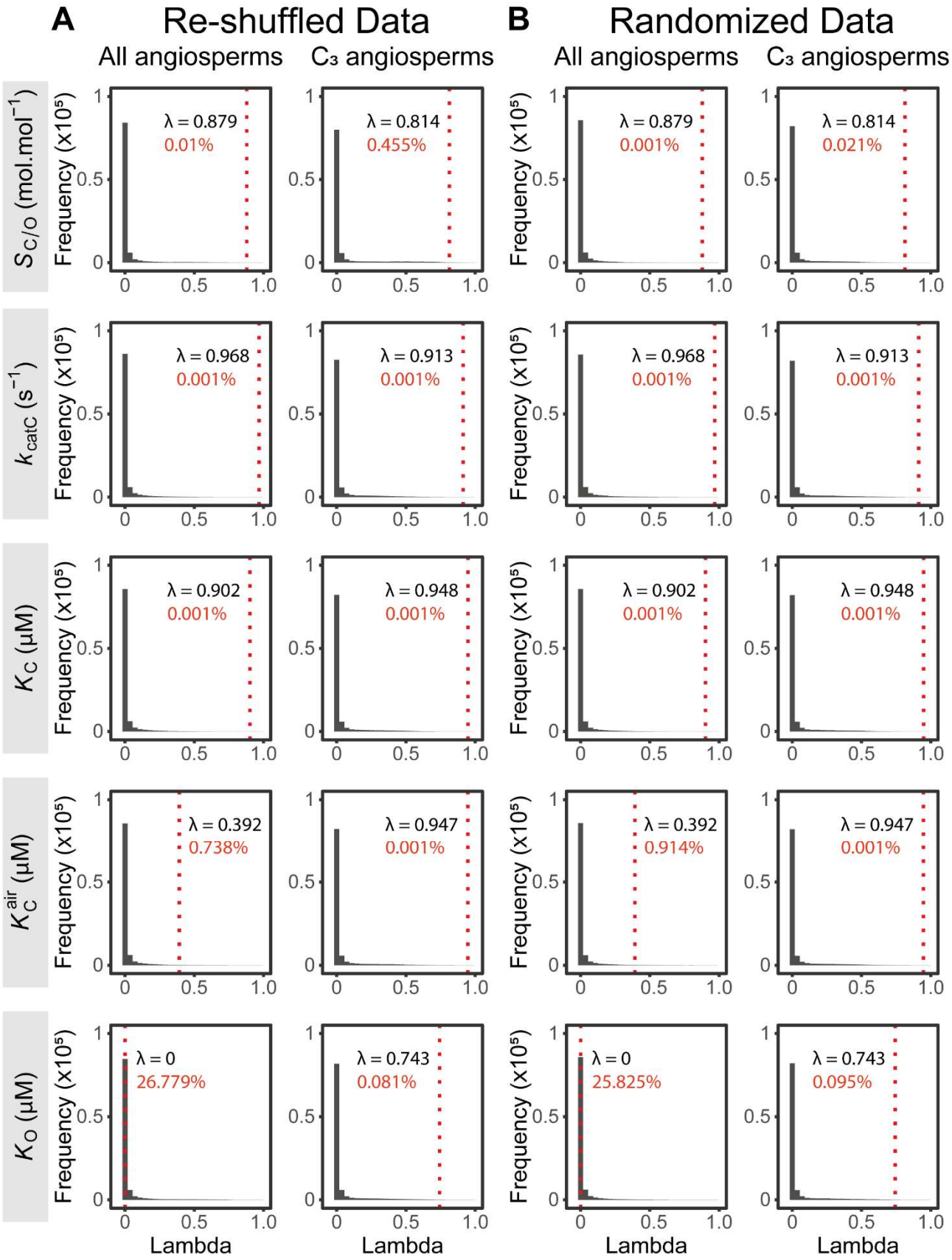
However, the presence of phylogenetic signal in Bouvier et al., (2021) was assessed using a combination of robust statistical methods based on both non-parametric permutation tests (Blomberg's *K*, Blomberg's *K**, Moran's *I* and Abouheif's *Cmean*) as well as likelihood ratio tests (in the utilized implementation of Pagel's lambda). In brief, permutation tests evaluate the probability that the null hypothesis of absence of signal is true by comparing the estimated test statistic for a given dataset to the distribution of values of this statistic obtained when randomly re-shuffling the data among species on the tree across a set number of replicates (n = 10,000 in Bouvier et al., (2021)). Conversely, likelihood ratio tests evaluate the probability that the null hypothesis is true by comparing the ratio of the log likelihood of the null hypothesis to that of the alternate hypothesis using the formula -2[log likelihood (null hypothesis) – log likelihood (alternative hypothesis)]. In both cases, if the computed p-value satisfies the given threshold of significance, the null hypothesis is rejected, and phylogenetic signal is deemed to be statistically significant. Specifically, in Bouvier et al., (2021), the majority of p-values reported for phylogenetic signal in rubisco kinetic traits were below p = 0.001 across the suite of detection methods used (Appendix A, table S2). This means that in most cases, there is less than a 0.1% chance of observing a similar or stronger phylogenetic signal by chance. Thus, it is not true that an arbitrary trait can spuriously generate a phylogenetic signal in a reliable manner. In fact, by definition, an arbitrary trait would produce a similar signal in fewer than 1 in 1,000 simulations. Nevertheless, although these statistical methods are powerful and perform well under a range of scenarios, we reaffirm below using two independent analyses that an arbitrary trait cannot reliably produce a phylogenetic signal. Thus, the above statement on this matter made by Tcherkez and Farquhar (2021) is mistaken.

First, to confirm the validity of our implemented phylogenetic tests and demonstrate that arbitrary traits cannot produce a statistically significant phylogenetic signal, we subject 100,000 replicates of a randomized assignment of kinetic trait data across species on the rubisco tree to phylogenetic signal analysis (i.e., recapitulating the methods of the permutation-based signal tests manually). Here, when using a cut-off threshold of significance at $p \leq 0.05$, interrogation of these randomized data revealed a significant signal could only be detected in 5% of the re-shuffled kinetic trait data when considering the majority of signal detection methods (i.e., exactly matching the proportion that

would be predicted to occur by chance), except Pagel's lambda which exhibited a considerably lower significance rate (Appendix A [and ***Supplemental File 1 online***], table S3). This result was similarly observed in the analysis across all angiosperms, as well as across the subset of C_3 angiosperms (Appendix A [and ***Supplemental File 1 online***], table S3). In total, this resulted in a type I (false positive) error rate in the randomized data at between 1.1 and 1.7% when taking the consensus result across at least 3 out of 5 phylogenetic signal detection methods (Appendix A [and ***Supplemental File 1 online***], table S3).

To further demonstrate that any arbitrary trait cannot generate a phylogenetic signal, 100,000 datasets were also randomly simulated *de novo* and subject to phylogenetic signal interrogation as above. To achieve this, we used the rubisco phylogenetic tree in Bouvier et al., (2021) and replaced the kinetic data for each species with randomly generated data. For this purpose, the mean and standard deviation of the real kinetic traits were used to guide the simulated data such that 100,000 arbitrary variables were produced with the same distributional properties as each trait. Analogous to above, when using a cut-off threshold of significance at $p \leq 0.05$, we observe a phylogenetic signal in 5% of all arbitrary traits in both the analysis across all angiosperms, as well as across the subset of C_3 angiosperms for the majority of detection methods (i.e., again corresponding to that which would be expected by chance) (Appendix A [and ***Supplemental File 1 online***], table S4). Accordingly, as above, the type I false positive rate at this cut-off threshold was observed at between 1.4 and 1.6% when considering the majority result in at least 3 phylogenetic signal detection methods (Appendix A [and ***Supplemental File 1 online***], table S4).

Thus, in summary, although it is possible to generate an artefactual phylogenetic signal using a randomly simulated trait, this scenario is exceptionally unlikely. Moreover, the strength of the phylogenetic signal observed among all simulated arbitrary variables is low (Appendix A, table S2; Appendix A [and ***Supplemental File 1 online***], table S3 – S5) and considerably weaker than that reported in rubisco kinetic traits (Bouvier et al., 2021). For example, when considering Pagel's lambda, the metric used to correct for phylogenetic effects in our downstream phylogenetic least squares regression analysis (Bouvier et al., 2021), on average fewer than 1 in 1,000 of the re-shuffled



Appendix A, Figure S3. The phylogenetic signal strength in randomly simulated arbitrary trait data. A) Frequency distribution plots to show the phylogenetic signal assessed by Pagel's lambda (λ) across 100,000 stochastically simulated datasets generated by randomly re-shuffling rubisco kinetic trait values across species on the phylogenetic tree. Each row represents data from a different

rubisco kinetic trait. Data are shown for the analysis of all angiosperms and C_3 angiosperms, respectively. The corresponding Pagel's lambda value measured in the real rubisco kinetic trait data is included in each plot and is visualised by the red vertical dotted line. The percentage of arbitrary simulations which produced an equivalent or stronger phylogenetic signal than that observed for the real rubisco kinetic trait is also included in each plot. B) As in (A), but for 100,000 stochastically simulated trait datasets generated by randomly sampling from normal distributions inferred from the mean and standard deviation of each real rubisco kinetic trait. In both (A) and (B) it should be noted that the vast majority of randomly simulated arbitrary data produce a very weak phylogenetic signal which approaches 0, irrespective of the kinetic trait or the method of random arbitrary data simulation utilized. The average phylogenetic signal strength and percentage of significant tests in the randomly simulated arbitrary trait data based on Pagel's lambda as well as the other signal detection methods utilized in the present study can be found in Appendix A [and ***Supplemental File 1 online***], table S3 and S4. The corresponding percentage of stochastically simulated datasets which produced an equivalent or stronger signal than that observed in the real rubisco kinetic traits, but inferred using the other phylogenetic signal detection methods can be found in Appendix A [and ***Supplemental File 1 online***], table S5. **[*End of Figure Legend*].**

---

data or the randomly simulated data produced a signal as strong or stronger then was observed for the real rubisco kinetic data (excluding K_O which had no phylogenetic signal in the analysis across all angiosperms) (Appendix A, Figure S3A and S3B; Appendix A [and ***Supplemental File 1 online***], table S5). This result corresponds with the p-values reported for phylogenetic signal in the rubisco kinetic traits in our original analysis and thus confirms the robustness of the phylogenetic signal detection methods utilized in Bouvier et al., (2021). Thus, in contrast to the claim of Tcherkez and Farquhar (2021), the probability that randomly distributed data would produce a significant phylogenetic signal is 1%, and the probability that this randomly distributed data would produce a signal comparable or stronger to that observed for real rubisco kinetics is less than 0.1%. This raises questions about the method used by Tcherkez and Farquhar (2021) to compute their significant phylogenetic signal in their randomly simulated trait given that no formal description or raw data was provided.

A.3.3 A phylogenetic signal is a universal feature of arbitrary traits distributed according to Brownian motion

In their Opinion article, Tcherkez and Farquhar (2021) further claimed that the phylogenetic signal in rubisco kinetics is artefactual because they can also observe the presence of a signal in a simulated trait that has been modelled to evolve on the rubisco tree by Brownian motion. However, this assertion demonstrates some misunderstanding about the principles of phylogenetic data interrogation. To explain, in the context of comparative biology, Brownian motion is a widely used model of evolution in which a trait is simulated to evolve stochastically in both direction and magnitude through time. Specifically, in this model, traits evolve by accruing incremental changes along the branches of a phylogenetic tree (i.e., from the last common ancestor at the root to all extant species at the terminal tips) by sampling at each node from a random distribution of possible changes with zero mean and finite constant variance. In this way, Brownian motion is assumed to represent a tractable model to approximate how trait evolution might occur in the real world under a wide range of scenarios. This is because simulated variables acquire several inherent properties which parallel those typically observed during the evolution of many real biological traits. 1) Trait variation among extant species conforms to a normal distribution with a mean equal to the ancestral trait value (i.e., trait evolution is non-directional). 2) Trait variation among extant species increases as a function of evolutionary time encapsulated by the tree (i.e., trait evolution is additive). 3) Trait variation among extant species co-varies as a function of shared ancestry (i.e., traits are strongly phylogenetically structured). Given this latter property, it follows that phylogenetic signal will inevitably be a universal feature of all traits simulated on a tree using Brownian motion, irrespective of the underlying phylogenetic tree on which they are simulated. To illustrate this point, we have assessed the presence of phylogenetic signal in 100 arbitrary BM traits simulated to have evolved on each of 100,000 simulated trees and demonstrate that a strong significant phylogenetic signal is observed in ~100% of cases (Appendix A [and ***Supplemental File 1 online***], table S6). This result is recapitulated if the number of terminal tips on the simulated trees are set to equal either the full number of angiosperms in the rubisco dataset, or the reduced number which perform C_3 photosynthesis (Appendix A [and ***Supplemental File 1 online***], table S6). Thus, the presence of

phylogenetic signal in traits simulated by Brownian motion on a tree is expected. Moreover, the fact that Tcherkez and Farquhar (2021) found phylogenetic signal in a simulated trait that was guaranteed to have strong phylogenetic signal due to the manner in which it was simulated does not contradict the finding that rubisco kinetic traits have phylogenetic signal, nor does it negate the requirement to account for this phylogenetic signal in downstream statistical analysis of the kinetic trait data. The "serious statistical problem" in rubisco kinetics that exists is not caused by random effects in the data and needs to be correctly considered when computing trait correlations on a phylogenetic tree.

A.3.4 Phylogenetic signal is not an artefact of biases in species sampling

Another criticism made by Tcherkez and Farquhar (2021) asserts that the presence of phylogenetic signal in rubisco kinetic traits is caused by biases in species sampling. Specifically, they argue that because groups of closely related species are present in the rubisco dataset, this has overestimated the phylogenetic signal in rubisco kinetic traits due to the effect of short branch lengths separating kinetically similar rubisco. However, as above, this argument reveals a misunderstanding about the principles of phylogenetic data interrogation. For instance, phylogenetic signal detection methods work by computing the pairwise co-variation between trait similarity and phylogenetic distance across species. If nearly interchangeable rubisco kinetics are present in a dataset between closely related sister species, this would be accounted for in the phylogenetic context of their recent ancestry and short branch lengths, not overestimated by this effect (Wiens et al., 2008). In fact, the central premise of phylogenetic comparative methods is that they are robust to species sampling, whereas non-phylogenetic methods are not. To illustrate this point, we have removed the species hypothesized to be problematic by Tcherkez and Farquhar (2021) from the analysis of all angiosperms and C_3 angiosperms (including plants in the *Oryza* clade, as well as those belonging to both the *Aegilops* and closely related *Triticum* clades which together share a convoluted evolutionary history (Petersen et al., 2006)) and have shown that an analogous phylogenetic signal is observed in this analysis (Appendix A [and ***Supplemental File 1 online***], table S7) as compared to in the original analysis when these species were present (Appendix A, table S2). This conclusion is true whether all angiosperms, or whether only C_3 angiosperms are considered (Appendix A, table S2; Appendix A

[and ***Supplemental File 1 online***], table S7). In summary, therefore, biases in species sampling including short branches separating nearly identical sister species, and overrepresentation of certain groups, are not responsible for the phylogenetic signal in rubisco kinetic traits. Indeed, it has in fact been shown that short branches have the inverse effect of causing phylogenetic signal to be underestimated due to complications arising with difficulties in resolving the correct phylogenetic history of species (Wiens et al., 2008).

A.3.5 Phylogenetic signal in rubisco kinetics is not an artefact of using rbcL-based trees

A further issue raised by Tcherkez and Farquhar (2021) is associated with the use of the *rbcL* gene to reconstruct the phylogenetic tree of species in our analysis. Specifically, they contend that tree inference based on *rbcL* sequence similarity across species may have caused inflation of phylogenetic signal due to overfitting. This is because mutations in *rbcL* which contribute to kinetic variation among species would also affect tree topology. However, we specifically addressed this possibility in our original study (Bouvier et al., 2021) and have previously shown this to not be an issue as we were able to replicate our results when using a phylogeny generated from only *rbcL* codon positions which contained synonymous substitutions across species (i.e., where columns in the alignment containing non-synonymous substitutions that could affect enzyme kinetics were removed) (Bouvier et al., 2021). Given that the use of this tree did not impact our analysis compared to that based on the complete *rbcL* tree, the phylogenetic signal in rubisco kinetics was confirmed to not be attributable to overfitting (Bouvier et al., 2021). As such, the full alignment *rbcL* tree was deemed appropriate for our analyses given that it has a long history in phylogenetic inference of species relationships (APG, 2016, 1998; Gielly and Taberlet, 1994) and more accurately reflects the evolutionary history of rubisco compared to the partial *rbcL* tree inferred from only synonymous substitutions.

Owing to limited availability of publicly sequenced nuclear (n = 26) or chloroplast (n = 8) genomes for species in the kinetic dataset, it was not possible to use whole genome phylogenomic approaches to infer the tree of species in our analysis. Nevertheless, to provide further reassurance in the present study that the phylogenetic signal in rubisco kinetics is not a consequence of the use of *rbcL* for

phylogenetic inference, we repeated our analysis using phylogenies inferred from the coding sequences of several other chloroplast encoded genes which are frequently utilized in systematics for species classification (Bohs and Olmstead, 1997; Ferguson, 1998; Hilu and Liang, 1997; Koch et al., 2001; Savolainen et al., 2000; Wolf, 1997; Yen and Olmstead, 2000) and importantly are unrelated to the kinetic traits of interest. Specifically, genes employed to infer these respective alternative trees include the ATP synthase beta subunit (*atpB*), the NADH dehydrogenase F (*ndhF*) subunit and maturaseK (*matK*) which was specifically argued by Tcherkez and Farquhar (2021) to reflect the evolutionary relationships of the C_4 species more accurately in our analysis compared to those inferred by *rbcL*. Analogous measurements of phylogenetic signal to those reported in our original paper are observed when these alternative gene trees are used (Appendix A, table S2; Appendix A [and ***Supplemental File 1 online***], table S8). In summary, therefore, the presence of phylogenetic signal in rubisco kinetics is not an artefact of using *rbcL*-based trees to infer phylogeny.

A.3.6 Phylogenetic signal is not an artefact of method-to-method variability in trait measurements

A separate concern raised by Tcherkez and Farquhar (2021) is related to the fact that our kinetic data was compiled from numerous sources in the literature. Specifically, in the analysis of angiosperms, the kinetic dataset was compiled across ten independent studies, including (Galmés et al., 2014; Kubien et al., 2008; Occhialini et al., 2016; Orr et al., 2016; Prins et al., 2016; Savir et al., 2010; Sharwood et al., 2016; von Caemmerer et al., 1994; Whitney et al., 2011; Zhu et al., 1998). Given that laboratory-to-laboratory variability exists in kinetic measurements due to inconsistencies in experimental protocols and techniques (Iñiguez et al., 2021), it was argued that such biases could confound the results of a phylogenetic signal analysis. Theoretically, this effect would be further exacerbated if studies have characterized rubisco among closely related clusters of species on the tree of life, which would thus have more comparable values to one another, but less comparable values to other distantly related species characterized in separate laboratories.

Although we agree that methodological-induced systematic biases are a genuine concern in analysis of any metadata, we have previously taken every effort to mitigate these effects from influencing our analysis. For example, the rubisco meta-dataset originally assembled by Flamholz et al., (2019) was

limited to include only measurements assayed under conditions of 25 °C and pH 7.8–8.0. Moreover, rubisco $S_{C/O}$ values in Bouvier et al., (2021) were normalised between studies using an oxygen electrode assay (Parry et al., 1989) and those using the improved high precision gas-phase-controlled ^{3}H-RuBP-fixation assay (Kane et al., 1994) by applying corrections based on wheat as an internal standard in both methods. Finally, study-specific errors were also limited in Bouvier et al., (2021) by taking averages across rubisco kinetics in the same species which were measured independently by different authors. Therefore, although it is impossible to completely avoid measurement variability across published studies, we have been pragmatic to lessen this impact where possible by adopting caution in data compilation (Flamholz et al., 2019) as well as in subsequent data processing (Bouvier et al., 2021).

Nevertheless, to confirm the robustness of our results to inter-study biases in measurements, we have repeated our analysis here using data derived from a single study in the meta-dataset so that there can be no influence of method-to-method or laboratory-to-laboratory variability. For this purpose, the subset of rubisco characterized in Orr et al., (2016) were analyzed, as it had sufficient species sampling for a robust statistical analysis. Phylogenetic interrogation of this dataset revealed a similar magnitude and significance of phylogenetic signal in rubisco (Appendix A [and ***Supplemental File 1 online***], table S9) compared to that which was originally reported from the meta-dataset (Appendix A, table S2). This same outcome is observed irrespective of whether all angiosperms, or exclusively C_3 angiosperms, were analysed (Appendix A, table S2; Appendix A [and ***Supplemental File 1 online***], table S9). Therefore, phylogenetic signal in rubisco kinetics is not an artefact of method-to-method or laboratory-to-laboratory variability in trait measurements.

In addition to the general concerns with our meta-analysis that are addressed above, a more specific criticism of method-to-method biases in our dataset was associated with the *Limonium* tribe. In particular, Tcherkez and Farquhar (2021) argue that the $S_{C/O}$ values of *Limonium* originally measured by (Galmés et al., 2014) as found in our dataset have been overestimated by ~13 mol mol^{-1} as compared to all other measured rubisco. Accordingly, they have claimed that this bias was the cause of the phylogenetic signal in this trait in our study. However, in the present analysis when such a correction of $S_{C/O}$ values in *Limonium* is carried out, we find that this does not cause an appreciable

difference in the magnitude of the phylogenetic signal in this trait either across all angiosperms, or across C_3 angiosperms (Appendix A [and ***Supplemental File 1 online***], table S10), as compared to the original analyses based on uncorrected values (Appendix A, table S2). An analogous conclusion is similarly obtained across both the analysis of all angiosperms and C_3 angiosperms if species in the *Limonium* clade were removed completely (Appendix A, table S2; Appendix A [and ***Supplemental File 1 online***], table S11). In combination, therefore, phylogenetic signal in rubisco kinetic traits is genuine and is not an artefact of the sampled species nor the data being compiled from different sources.

A.3.7 Phylogenetic signal is not an artefact of C_3 vs C_4 photosynthesis

Given our analysis is robust to all criticisms made by Tcherkez and Farquhar (2021), an interesting question of fundamental importance is why strong pervasive phylogenetic signal exists in rubisco kinetics (Bouvier et al., 2021). On this matter, despite their objections, Tcherkez and Farquhar (2021) postulate that this phenomenon is primarily driven by differences in the biochemical niche of the enzyme across the tree of life. Specifically, given the widespread kinetic differences associated with adaptation to intracellular CO_2 levels in plants with C_3 (higher CO_2 specificity) and C_4 metabolisms (higher CO_2 turnover) (Bouvier et al., 2021), Tcherkez and Farquhar (2021) claim that C_3 vs. C_4 kinetic differences are responsible for the observed phylogenetic signal.

However, the above hypothesis put forward by Tcherkez and Farquhar (2021) is flawed by confusion between global phylogenetic signal (referred to here and in Bouvier et al., (2021) as simply “phylogenetic signal”) and local phylogenetic signal (analysed in Tcherkez and Farquhar, (2021), but not discussed in Bouvier et al., (2021) or in the present study until this point). In brief, local measures of phylogenetic signal, such as local Moran’s *I* (Anselin, 1995), serve to identify unique and interesting patterns of trait variation in a group of species which are not conserved in neighbouring clades (i.e., the kinetic outliers of rubisco in C_4 compared to C_3 sister species along the angiosperm tree). In contrast, global phylogenetic signal as assessed here and by Bouvier et al., (2021) is different and instead characterises the average statistical co-variation between trait values and phylogenetic distances across all pairwise combinations of species on the tree. As such, global metrics of phylogenetic signal (including Pagel’s lambda (Pagel, 1999), Blomberg’s *K* and *K**

(Blomberg et al., 2003), Moran's *I* (Gittleman and Kot, 1990) and Abouheif's *Cmean* (Abouheif, 1999)) are ultimately used to correct for the phylogenetic non-independence of data in phylogenetic comparative models. Based on this distinction, although kinetic changes associated with transition to C_4 photosynthesis in discrete pockets on the phylogenetic tree will produce hotspots of local phylogenetic signal, this effect will reduce the estimated global phylogenetic signal which we measure. This is because the convergence and homoplasy of the C_4 trait weakens the average statistical dependence between trait similarity and phylogenetic relatedness among the group of sampled species as a whole (Hansen and Martins, 1996; Kamilar and Cooper, 2013). Indeed, this fact was shown in the results of our original study (Bouvier et al., 2021) as well as those presented here (Appendix A, table S2; Appendix A [and ***Supplemental File 1 online***], table S6 – S11) whereby inclusion of C_4 and other C_3-C_4 intermediary species reduces the signal strength relative to when only C_3 angiosperms are considered. Moreover, this result is also observed in the analysis presented by Tcherkez and Farquhar (2021), although this effect is misinterpreted by these authors. Specifically, using k_{catC} as an example, Tcherkez and Farquhar (2021) show that C_4 angiosperms are associated with local phylogenetic signal when all species are considered, however they detect no global phylogenetic signal in this analysis (Figure 2 of Tcherkez and Farquhar, (2021)). In contrast, when C_4 and other intermediary C_3-C_4 and C_4-like species are ignored, Tcherkez and Farquhar (2021) show the local phylogenetic signal associated with these groups is lost and global phylogenetic signal accordingly becomes significant at short phylogenetic distances of the remaining species (Figure 2 of Tcherkez and Farquhar, (2021)). This result demonstrates that although C_3 vs. C_4 kinetic differences drive local phylogenetic signal, they are not responsible (and in fact weaken) the global phylogenetic signal which is of importance to the discussion of Bouvier et al., (2021). This issue was already discussed in our original study (Bouvier et al., 2021). Thus, the contention of Tcherkez and Farquhar (2021) that C_4-mediated kinetic changes are responsible for the observed phylogenetic signal across the whole tree is incorrect, given that this signal is detected when analysing only C_3 angiosperms (as originally reported (Bouvier et al., 2021)), and the strength of the signal in the analysis of C_3 angiosperms is greater than when both C_3 and C_4 angiosperms are considered together.

A.4 Discussion

Rubisco (ribulose-1,5-bisphosphate [RuBP] carboxylase/oxygenase) is the primary entry point for carbon into the biosphere (Field et al., 1998; Tabita et al., 2008) and accordingly, is the source of almost all organic carbon which has ever existed. As such, gaining an appreciation of how rubisco has evolved is of fundamental importance to our understanding of life on our planet. Interestingly, however, rubisco has long presented an evolutionary paradox. This is because, despite the central importance of rubisco in underpinning host autotroph metabolism and ultimately the global food chain, the enzyme is considered by many to be an inefficient catalyst under the high CO_2, low O_2 conditions of the present day. This is due to the fact that rubisco has a modest turnover rate for its primary CO_2 reaction (Badger et al., 1998; Bar-Even et al., 2011) and also catalyses a costly secondary reaction with O_2 that culminates in the loss of previously fixed carbon (Bowes et al., 1971; Chollet, 1977; Sharkey, 2020). Although an updated examination has suggested that rubisco's kinetics are overall perhaps not as bad as often assumed (Bathellier et al., 2018), these puzzling catalytic limitations of the enzyme have nevertheless caused rubisco to be an enigma and the source of intensive and sustained research interest for over 40 years.

Until our phylogenetically resolved investigation of rubisco kinetic evolution (Bouvier et al., 2021), the principal hypothesis forwarded to explain the above "rubisco paradox" proposed that severe catalytic trade-offs exist between the kinetic traits of the enzyme. Specifically, this theory was first pioneered by two cross-species analyses (Savir et al., 2010; Tcherkez et al., 2006) which reported strong antagonistic correlations between rubisco kinetic traits and posited that these correlations were caused by unavoidable chemical constraints on its catalytic mechanism (despite a limited understanding of this chemical mechanism). In this way, these kinetic trait interdependencies were understood to provide a ceiling on rubisco optimisation by limiting the adaptative capacity of its individual kinetic traits (Savir et al., 2010; Tcherkez et al., 2006). However, both these analyses which first described the rubisco kinetic trait trade-offs (Savir et al., 2010; Tcherkez et al., 2006), in addition to all subsequent analyses which have re-investigated these trade-offs among different taxa (Flamholz et al., 2019; Iñiguez et al., 2020; Young et al., 2016), did not account for the phylogenetic

context of the sampled species being considered in the analysis. This is a serious statistical problem because species (and the rubisco they encode) cannot be treated as independent observations in comparative studies owing to the shared ancestry of all organisms on the hierarchical tree of life (Felsenstein, 1985). Thus, not accounting for phylogenetic non-independence has potentially violated one of the core assumptions of the conventional statistical methods which were applied (i.e., independence of residuals; Appendix A, table S1; Figure S1) meaning that the results drawn from these analyses may possibly be, wholly or in part, artefactual.

To assess whether the methodological flaws of previous cross-species analyses has led to mistaken inferences about rubisco, in Bouvier et al., (2021) we re-investigated the kinetic evolution of this enzyme in a phylogenetically resolved manner. We discovered for the first time that rubisco kinetic traits exhibit strong and significant phylogenetic signal (Bouvier et al., 2021). The presence of this signal means that kinetic trait values are statistically similar across enzymes as a function of their shared evolutionary history (i.e., the kinetic traits exhibit strong phylogenetic non-independence). Accordingly, we re-evaluated the correlations between rubisco kinetic traits when accounting for this phylogenetic signal and found that only canonical trade-offs between the Michaelis constant for CO_2 (K_C) and carboxylase turnover (k_{catC}), and between the Michaelis constants for CO_2 and O_2 (K_O), were robust to phylogenetic effects (Appendix A, Figure S2A – S2C) (Bouvier et al., 2021). Though, the strength of these trade-offs were moderate and considerably attenuated relative to earlier estimates (Appendix A, Figure S2A – S2C) (Bouvier et al., 2021). Therefore, although catalytic trade-offs exist and constrain rubisco adaptation, these are not absolute, and have been previously overestimated due to phylogenetic biases in the kinetic data (Appendix A, Figure S2A – S2C). This means that kinetic traits have been able to evolve largely independently of one another during the diversification of rubisco. Instead, when we further investigated the constraints on the enzyme, we found that phylogenetic constraints explained more variation in rubisco kinetics, and have thus had a larger impact on limiting enzyme adaptation, compared to the combined action of all catalytic trade-offs (Bouvier et al., 2021). Thus, although rubisco catalytic trade-offs exist, these represent a less serious constraint on rubisco kinetic adaptation than previously thought. In contrast, phylogenetic constraints, likely caused by slow molecular evolution in rubisco (Bouvier et al., 2022) and more

general constraints on the molecular evolution of chloroplast encoded genes (Robbins and Kelly, 2022), have presented a more significant barrier to improved rubisco catalytic efficiency.

In Bouvier et al., (2021) we made every effort to carefully describe the method by which we have computed our phylogenetic signal in rubisco kinetic data. In addition, we were also conscious to discuss what this phylogenetic signal means, as well as to explain in detail how the presence of this signal has incorrectly resulted in overestimated kinetic trait correlation coefficients in previous analyses. However, our results have been brought under criticism by Tcherkez and Farquhar, (2021). Specifically, these authors have argued that the phylogenetic signal we detect in rubisco kinetics is a consequence of computational artefacts due to a combination of biases associated with species sampling (including the use of near identical sister species as well as the overrepresentation of certain groups), the use of *rbcL*-based trees for phylogenetic inference, method-to-method and laboratory-to-laboratory variability among kinetic measurements in the rubisco meta-dataset, and finally, homoplasy of the C_4 trait across sampled species. On this matter, Tcherkez and Farquhar (2021) argue that our phylogenetic signal in rubisco kinetics is not valid because they can also generate a signal using our method with both a randomly simulated arbitrary trait as well as a trait distributed according to Brownian motion. In the present response, we critically review each of these claims on a point-by-point basis by re-visiting and extending our original analysis. In doing so, we provide clear evidence that all of the criticisms raised by Tcherkez and Farquhar (2021) were either incorrect, arose from a misunderstanding of phylogenetic comparative methods, or were already clearly addressed in our original manuscript and were overlooked by the authors. As such, we reaffirm that our results in Bouvier et al., (2021) are robust. Thus, our original conclusions, as summarized above, are unaltered.

To avoid any confusion on the points made by Tcherkez and Farquhar, (2021), it is worthwhile discussing further the subject of C_3 vs. C_4 photosynthesis in the context of measuring phylogenetic signal in rubisco kinetics. As described by our previous results (Bouvier et al., 2021) and those in (Capó-Bauçà et al., 2022; Cummins, 2021; Iñiguez et al., 2020), we agree that kinetic adaptations in rubisco occur in conjunction with the emergence of CO_2-concentrating mechanisms. For example, rubisco evolve to become faster (increased carboxylase turnover) and less specific for CO_2

(decreased specificity and CO_2 affinity) during transition from C_3 to C_4 photosynthesis (Bouvier et al., 2021). Nevertheless, although co-evolution between rubisco and C_4 photosynthesis is apparent, as underpinned by changes in the gaseous micro-environment of the enzyme, this is not responsible for the phylogenetic signal in rubisco kinetic traits as proposed by Tcherkez and Farquhar (2021). Indeed, rubisco kinetic adaptation associated with the convergent evolution of C_4 photosynthesis in discrete clusters of species on the phylogenetic tree in fact has the opposite effect of degrading the pairwise statistical association between phylogenetic distance and kinetic distance across all sampled rubisco as a whole. Thus, although differences in prevailing gaseous conditions around rubisco exist between C_3 and C_4 species and underpin differences in the trajectory of rubisco adaptation, this does not drive (and in fact weakens) the observed phylogenetic signal in rubisco kinetics.

Although the present work is focussed on angiosperms, it should be noted that we have previously observed identical results in rubisco across the tree of life (Appendix A, table S2) (Bouvier et al., 2021). Thus, although rubisco evolution is constrained by catalytic trade-offs, it is instead more limited by phylogenetic constraint. We propose that this phylogenetic constraint arises from a combination of a high degree of purifying selection (Robbins and Kelly, 2022), the requirement for high levels of transcript and protein abundance (Kelly, 2018; Robbins and Kelly, 2022; Seward and Kelly, 2018), the requirement for maintaining complementarity to a wide array of molecular chaperones which assist in protein folding and assembly (e.g., Raf1, Raf2, RbcX, BSD2, Cpn60/Cpn20) and metabolic regulation (e.g., rubisco activase) (Aigner et al., 2017; Carmo-Silva et al., 2015), and finally, the need to preserve overall protein stability within the molecular activity-stability trade-offs (Cummins et al., 2018; Duraõ et al., 2015; Studer et al., 2014). These factors combined contribute to the exceedingly slow rate of molecular evolution in *rbcL* (Bouvier et al., 2022) which presents a major barrier on rubisco optimisation.

A.5 Conclusions

Prior to our phylogenetically resolved investigation of rubisco evolution (Bouvier et al., 2021), previous studies that measured correlations between rubisco kinetic traits were computed in the

absence of accounting for the phylogenetic non-independence of the data. We described how this omission has led to the mistaken inference that the biochemical landscape of rubisco is severely constrained by catalytic trade-offs (Bouvier et al., 2021). In the present study, we revisit these analyses to address a series of criticisms of our work that were put forward in a recent opinion article (Tcherkez and Farquhar, 2021). Here, we demonstrate that all of these criticisms were unfounded, and thus our original conclusions remain unaltered. Namely, strong phylogenetic signal exists in rubisco kinetic traits and causes an overestimation of catalytic trade-offs unless correctly accounted for. In actual fact, phylogenetic constraints have limited rubisco kinetic adaptation to a greater extent than the combined action of catalytic trade-offs. These phylogenetic constraints are caused by multiple evolutionary factors that act to limit rubisco molecular sequence evolution (Bouvier et al., 2022) and combined, help to explain why the enzyme is poorly efficient under present-day conditions but better adapted to the former high CO_2, low O_2 atmosphere in which it evolved. These results agree with results from other updated cross-species analyses which have similarly emphasized that kinetic trait correlations are generally too weak to support the rubisco trade-off model (Cummins et al., 2019, 2018; Flamholz et al., 2019; Galmés et al., 2014). Moreover, these conclusions are also consistent with findings from recent studies which have characterized novel enzyme variants in both nature (e.g., Aguiló-Nicolau et al., 2023) and in rubisco engineering experiments (Wilson et al., 2018; Zhou and Whitney, 2019) that deviate from the previously proposed catalytic trade-offs.

A.6 Materials and Methods

A.6.1 Kinetic data

The rubisco kinetic dataset used in this work was compiled across various sources in the primary literature (Bouvier et al., 2021; Flamholz et al., 2019). This data includes experimentally determined measurements of wild-type rubisco assayed under conditions of pH 7.8–8.0 and 25 °C for CO_2/O_2 specificity ($S_{C/O}$), maximum carboxylase turnover rate per active site (k_{catC}), and the respective Michaelis constant (i.e., the substrate concentration at half saturated catalysed rate) for both CO_2 (K_C) and O_2 (K_O) substrates. In addition, measurements of the Michaelis constant for CO_2 in 20.95% O_2 ambient air (K_C^{air}) were also available as previously derived from K_C and K_O using the formula K_C

+ (K_C [O_2]/K_O), where 20.95% [O_2] in water is 253 µM (Bouvier et al., 2021). As previously described (Bouvier et al., 2021), modifications of this dataset have been performed by normalizing between $S_{C/O}$ measurements that were determined using an oxygen electrode assay (Parry et al., 1989) and those determined using the high precision gas-phase-controlled ^{3}H-RuBP-fixation assay (Kane et al., 1994) by using wheat as an internal standard in both. In addition, this dataset was also previously modified to average across duplicate entries of species (including synonyms) between studies (Bouvier et al., 2021). It should be noted that in the present study, only data for the 137 angiosperm species with complete measurements of $S_{C/O}$, k_{catC}, K_C, K_C^{air} and K_O were considered given that this is the taxonomic group which was solely focused on by Tcherkez and Farquhar (2021). Further, as before, all kinetic traits were log transformed prior to interrogation to satisfy the distributional assumptions of the statistical tests being performed.

A.6.2 Phylogenetic tree inference based on rbcL and other gene sequences

All phylogenetic trees inferred from the coding sequence of the rubisco large subunit (*rbcL*) were obtained from our previous work (Bouvier et al., 2021). This includes the separate *rbcL* trees which describes the evolutionary history of all angiosperms, as well as the subset of angiosperms which perform C_3 photosynthesis. Phylogenetic trees based on other chloroplast encoded genes including maturaseK (*matK*), ATP synthase beta subunit (*atpB*), and the NADH dehydrogenase F (*ndhF*) subunit were generated in the present study following the exact method previously described (Bouvier et al., 2021). Specifically, coding sequences of *matK*, *atpB*, and *ndhF* were obtained from NCBI (https://www.ncbi.nlm.nih.gov/) for as many of the angiosperm species in the kinetic dataset as possible. Multiple sequence alignments were generated for each respective gene using mafft L-INS-i (Katoh and Standley, 2013), and alignments were processed to remove any partial, chimeric, or erroneously annotated sequences. Finally, bootstrapped maximum-likelihood phylogenetic trees were inferred from each non-gapped sequence alignment by IQ-TREE (Nguyen et al., 2015) using the ultrafast bootstrapping method with 1,000 replicates and the Shimodaira–Hasegawa approximate-likelihood ratio branch test, with the best fitting model of sequence evolution chosen automatically. All trees were rooted manually using Dendroscope (Huson and Scornavacca, 2012). For each of these inferred trees based on *matK*, *atpB*, and *ndhF* gene sequences, a corresponding

tree containing only C_3 species was obtained using the function drop.tip from the 'ape' package (Paradis et al., 2004; Paradis and Schliep, 2019) in the R environment.

Due to the differences in phylogenetic information available for tree building from the alignments of *matK*, *atpB*, and *ndhF* genes as compared to the alignment based on the *rbcL* gene, additional polytomies which include species with terminal zero-length branches were detected in these alternative chloroplast gene trees (due to species possessing 100% sequence identity for the gene in question). However, it was not appropriate to condense these zero-branch length nodes into single datapoints as previously performed in the analysis based on the *rbcL*-based trees (Bouvier et al., 2021). This is because these other chloroplast genes are unrelated to rubisco kinetic traits. Thus, averaging kinetic data across zero-branch length species on these other trees is incorrect in the context of the downstream phylogenetic analysis being performed, given that these nodes are known to possess differences in their rubisco sequences which are not captured by the tree. Thus, to get around this issue and maximize the number of species taken forward for analysis, zero-terminal length branches in *matK*, *atpB*, and *ndhF* trees were arbitrarily resolved into fully bifurcating trees by using the function multi2di in the 'ape' R package (Paradis et al., 2004; Paradis and Schliep, 2019) and increasing terminal branch length values by a minimal value of 0.000001. The total number of angiosperms and the total number of C_3 angiosperms encapsulated by each of these alternative gene trees can be found in (Appendix A [and ***Supplemental File 1 online***], table S8). These alternative trees are provided in Newick format in (***Supplemental File 2 online***).

A.6.3 Generation of randomly simulated arbitrary traits

To evaluate whether randomly simulated data could produce an artefactual phylogenetic signal, two separate approaches were employed. First, 100,000 random permutations of the data were performed manually such that in each permutation the kinetic trait data for the species set was subject to a Fisher-Yates shuffle. Each permutation was subject to phylogenetic signal analysis as described below and the proportion of permutations that obtained an equivalent or larger phylogenetic signal to that observed in the real kinetic trait data were recorded. Second, 100,000 stochastically simulated trait datasets were generated by randomly sampling from normal distributions inferred from the mean and standard deviation of each real rubisco kinetic trait. To

achieve this, the rnorm function was employed from the 'compositions' package (van den Boogaart and Tolosana-Delgado, 2008) in the R environment. Each simulation was subject to phylogenetic signal analysis as described below and the proportion of simulated trait dataset that obtained an equivalent or larger phylogenetic signal to that observed in the real kinetic data were recorded.

A.6.4 Simulation of kinetic data on phylogenetic trees by Brownian motion

To show that a phylogenetic signal is a universal feature of all traits which are distributed according to Brownian motion on any underlying phylogenetic tree, 100,000 bifurcating trees were randomly simulated using the rtree function from the 'ape' package (Paradis et al., 2004; Paradis and Schliep, 2019) in R. This process was repeated for trees containing both the same number of terminal tips as the full set of angiosperms in the rubisco kinetic dataset and the same number of tips as the subset of C_3 angiosperms, respectively. Next, for each tree, 100 quantitative traits were simulated to evolve under a standard Brownian motion model using the function fastBM in the R package 'phytools' (Revell, 2012). In total, this resulted in 10,000,000 unique Brownian motion trait simulations for the analysis of all angiosperms and C_3 angiosperms, respectively. Each simulated trait was subject to phylogenetic signal analysis using the methods outlined below.

A.6.5 Phylogenetic signal analysis

The magnitude and significance of phylogenetic signal in a given trait on a given underlying phylogenetic tree was computed using five independent detection methods, including Pagel's lambda (Pagel, 1999), Moran's I (Gittleman and Kot, 1990), Abouheif's Cmean (Abouheif, 1999) and Blomberg's K and K* (Blomberg et al., 2003). Specifically, these tests work by assessing the statistical association between trait and phylogenetic similarity among all pairwise combinations of species on the tree. Implementation of these phylogenetic signal detection tests was performed in the R environment as previously described (Bouvier et al., 2021).

A.7 Supplemental Tables

Appendix A, Box 1. A list of definitions for key terms as they are applied in the context of the present study and in our original study (Bouvier et al., 2021).

Term	Definition
Kinetic trait	A term used to refer to all constants and rates which describe the biophysiochemical properties of an enzyme reaction.
Phylogenetic signal	A statistical measure of co-variation between the evolutionary relatedness and phenotypic similarity among sampled species.
Phylogenetic non-independence	The inherent nested structure in trait data which follows the shared ancestry among sampled species on the phylogenetic tree.
Phylogenetic comparative method	A comparative statistical method which analyses cross-species data whilst accounting for the underlying evolutionary relationships among the sampled species on the phylogenetic tree.
Phylogenetic generalized least squares regression	A type of phylogenetic comparative method which analyses associations between two (or more) continuous variables. This model corrects for the phylogenetic signal in the residuals of the traits by attributing differential weights to species depending on their evolutionary relatedness to one another. This is also commonly referred to as a 'phylogenetic regression', 'phylogenetic generalised linear regression' and 'pgls'.
Residual error	An error term which denotes the vertical distance of a data point from its arithmetic mean in X~Y trait space (defined by the optimal line of best of fit to the data).
Catalytic trade-off	A mechanistic constraint on the chemical mechanism of an enzyme. In the case of rubisco, these trade-offs have been inferred from observations of antagonistic correlations between the enzyme's kinetic traits.
Phylogenetic constraint	The cohort of factors that constrain molecular evolution and thus, in turn, functional adaptation of a protein.

Appendix A, Table S1. Overview of the key differences between conventional non-phylogenetic generalized least squares regression models and phylogenetic generalised least squares regression models.

	Assumes independence?	Data Requirements
Non-phylogenetic regression	✓ Assumes datapoints are independent and randomly sampled	✓ Requires comparative trait data ✗ Does not require phylogenetic information (datapoints are assumed to be independent and not related in phylogenetic space)
Phylogenetic regression	✗ Does not assume independence among datapoints	✓ Requires comparative trait data ✓ Requires phylogenetic information to measure and correct for the phylogenetic signal (and phylogenetic non-independence) of species datapoints

Appendix A, Table S2. The phylogenetic signal strength and associated significance level in rubisco kinetic traits using five different signal detection methods, as reported in (Bouvier et al., 2021). Data are shown from the analysis of angiosperms, C_3 angiosperms, and all studied photosynthetic organisms, respectively.

NOTE. Statistics are rounded to three decimal places and significance values are represented as α levels, where α = 0.001 if P < 0.001, α = 0.01 if 0.001 < P < 0.01, α = 0.05 if 0.01 < P < 0.05, and α = ns if P > 0.05.

Data	*Kinetic*	*Cmean*		*I*		*K*		*K**		*Lambda*	
	Trait	*Stat*	*α*	*Stat*	*α*	*Stat*	*α*	*Stat*	*α*	*Stat*	*α*
All	$S_{C/O}$	0.516	0.001	0.425	0.001	0.001	ns	0.001	ns	0.879	0.001
Angiosperms	k_{catC}	0.350	0.001	0.224	0.001	0.003	0.001	0.003	0.001	0.968	0.001
	K_C	0.282	0.001	0.248	0.001	0.001	0.01	0.002	0.01	0.902	0.001
	K_C^{air}	0.234	0.001	0.242	0.001	0.001	0.05	0.001	0.05	0.392	ns
	K_O	0.032	ns	−0.070	ns	0	ns	0	ns	0	ns
C_3	$S_{C/O}$	0.533	0.001	0.453	0.001	0	ns	0.001	ns	0.814	0.001
Angiosperms	k_{catC}	0.387	0.001	0.234	0.001	0.002	0.01	0.002	0.01	0.913	0.001
	K_C	0.449	0.001	0.341	0.001	0.001	0.05	0.001	0.05	0.948	0.001
	K_C^{air}	0.398	0.001	0.317	0.001	0.001	0.05	0.002	0.01	0.947	0.001
	K_O	0.279	0.001	0.167	0.01	0	ns	0	ns	0.743	0.001
Photosynthetic Organisms	$S_{C/O}$	0.712	0.001	0.229	0.001	0.005	0.001	0.002	0.001	1.005	0.001
	k_{catC}	0.538	0.001	0.29	0.001	0.002	0.001	0.001	0.001	0.975	0.001
	K_C	0.705	0.001	0.277	0.001	0.002	0.001	0.001	0.001	0.948	0.001
	K_C^{air}	0.609	0.001	0.299	0.001	0.001	0.001	0.001	0.01	0.924	0.001
	K_O	0.170	0.001	0.004	ns	0	ns	0	ns	0.603	0.001

Appendix A, Table S3. The phylogenetic signal strength (mean ± S.E.) and percent of significant tests in arbitrary trait data generated by randomly re-shuffling real rubisco kinetic trait data across species on the phylogenetic tree using 100,000 replicates. Phylogenetic signal is measured using five signal detection methods. Data are shown from the analysis of all angiosperms and C_3 angiosperms, respectively.

Data	Arbitrary Trait	*Cmean*		*I*		*K*		*K**		*Lambda*		Consensus
		Stat	*% Sig*	*Stat*	*% Sig*	*Stat*	*% Sig*	*Stat*	*% Sig*	*Stat*	*% Sig*	*% Sig*
All Angiosperms	$S_{C/O}$	-0.008 ± 0	5.052	-0.008 ± 0	5.059	0 ± 0	5.165	0 ± 0	5.133	0.031 ± 0	1.817	1.075
	k_{catC}	-0.009 ± 0	5.019	-0.008 ± 0	4.994	0 ± 0	4.963	0 ± 0	5.066	0.017 ± 0	1.132	1.656
	K_C	-0.008 ± 0	5.080	-0.009 ± 0	4.898	0 ± 0	5.015	0 ± 0	4.92	0.018 ± 0	1.163	1.448
	K_C^{air}	-0.009 ± 0	4.882	-0.009 ± 0	4.951	0 ± 0	5.059	0 ± 0	5.097	0.018 ± 0	1.174	1.473
	K_O	-0.009 ± 0	4.945	-0.009 ± 0	4.902	0 ± 0	4.898	0 ± 0	4.946	0.026 ± 0	1.464	1.061
C_3 Angiosperms	$S_{C/O}$	-0.012 ± 0	5.000	-0.011 ± 0	5.086	0.001 ± 0	5.097	0.001 ± 0	5.053	0.056 ± 0	3.055	1.242
	k_{catC}	-0.012 ± 0	4.964	-0.011 ± 0	5.043	0 ± 0	5.023	0 ± 0	4.954	0.032 ± 0	1.419	1.625
	K_C	-0.012 ± 0	4.997	-0.012 ± 0	4.951	0 ± 0	5.054	0 ± 0	5.049	0.034 ± 0	1.521	1.460
	K_C^{air}	-0.011 ± 0	5.014	-0.011 ± 0	4.867	0 ± 0	5.013	0 ± 0	5.048	0.034 ± 0	1.596	1.439
	K_O	-0.011 ± 0	4.992	-0.011 ± 0	5.059	0 ± 0	5.018	0 ± 0	5.015	0.036 ± 0	1.671	1.485

Appendix A, Table S4. The phylogenetic signal strength (mean ± S.E.) and percent of significant tests in arbitrary trait data generated by randomly simulating data with the same distributional properties as each real rubisco kinetic trait using 100,000 replicates. Phylogenetic signal is measured using five signal detection methods. Data are shown from the analysis of all angiosperms and C_3 angiosperms, respectively.

Data	Arbitrary Trait	*Cmean*		*I*		*K*		*K**		*Lambda*		Consensus
		Stat	*% Sig*	*Stat*	*% Sig*	*Stat*	*% Sig*	*Stat*	*% Sig*	*Stat*	*% Sig*	*% Sig*
All Angiosperms	$S_{C/O}$	-0.009 ± 0	4.994	-0.008 ± 0	5.101	0 ± 0	4.912	0 ± 0	4.957	0.018 ± 0	1.232	1.515
	k_{catC}	-0.009 ± 0	5.050	-0.008 ± 0	4.931	0 ± 0	4.947	0 ± 0	4.915	0.018 ± 0	1.208	1.444
	K_C	-0.008 ± 0	5.100	-0.008 ± 0	5.160	0 ± 0	4.966	0 ± 0	5.040	0.019 ± 0	1.201	1.569
	K_C^{air}	-0.008 ± 0	5.098	-0.008 ± 0	5.117	0 ± 0	5.087	0 ± 0	5.114	0.018 ± 0	1.260	1.547
	K_O	-0.009 ± 0	4.941	-0.009 ± 0	4.904	0 ± 0	4.990	0 ± 0	4.973	0.018 ± 0	1.192	1.434
C_3 Angiosperms	$S_{C/O}$	-0.012 ± 0	4.957	-0.012 ± 0	4.991	0 ± 0	5.096	0 ± 0	5.004	0.035 ± 0	1.545	1.473
	k_{catC}	-0.011 ± 0	5.083	-0.011 ± 0	5.013	0 ± 0	4.903	0 ± 0	4.892	0.035 ± 0	1.607	1.463
	K_C	-0.011 ± 0	5.012	-0.011 ± 0	5.038	0 ± 0	5.065	0 ± 0	5.043	0.035 ± 0	1.542	1.483
	K_C^{air}	-0.011 ± 0	5.083	-0.011 ± 0	5.004	0 ± 0	5.031	0 ± 0	5.02	0.035 ± 0	1.644	1.567
	K_O	-0.012 ± 0	4.919	-0.012 ± 0	4.855	0 ± 0	5.004	0 ± 0	4.96	0.035 ± 0	1.567	1.435

Appendix A, Table S5. The percent of 100,000 arbitrary traits which exhibited a phylogenetic signal that was comparable or stronger than that observed in the real rubisco kinetic data. Arbitrary traits were generated by randomly re-shuffling real rubisco kinetic trait data across species on the phylogenetic tree (Appendix A, table S3) or by randomly simulating data with the same distributional properties as each real rubisco kinetic trait (Appendix A, table S4). Phylogenetic signal is measured using five signal detection methods. Data are shown from the analysis of all angiosperms and C_3 angiosperms, respectively.

Data	Arbitrary Trait	Re-shuffled					Simulated				
		Cmean	***I***	***K***	***K****	***Lambda***	***Cmean***	***I***	***K***	***K****	***Lambda***
All Angiosperms	$S_{C/O}$	0.001	0.001	7.087	8.279	0.010	0.001	0.001	2.980	3.613	0.001
	k_{catC}	0.001	0.013	0.005	0.004	0.001	0.001	0.023	0.006	0.005	0.001
	K_C	0.001	0.005	0.204	0.108	0.001	0.003	0.009	0.224	0.103	0.001
	K_C^{air}	0.022	0.007	4.740	3.056	0.738	0.018	0.004	4.674	2.998	0.914
	K_O		85.91	96.29	95.18			85.01	98.31	97.53	
		24.942	5	6	7	26.779	25.116	4	3	4	25.825
C_3 Angiosperms	$S_{C/O}$			40.11	38.53				29.32	26.75	
		0.001	0.001	8	0	0.455	0.001	0.001	6	8	0.021
	k_{catC}	0.001	0.033	0.434	0.331	0.001	0.001	0.045	0.512	0.384	0.001
	K_C	0.001	0.001	3.246	2.306	0.001	0.001	0.001	3.296	2.337	0.001
	K_C^{air}	0.001	0.001	1.423	0.965	0.001	0.001	0.001	1.304	0.869	0.001
	K_O			47.35	42.13				46.75	41.46	
		0.012	0.648	0	6	0.081	0.007	0.614	5	7	0.095

Appendix A, Table S6. The phylogenetic signal strength (mean ± S.E.) and percent of significant tests in arbitrary Brownian motion traits distributed across randomly simulated phylogenetic trees using 100 trait replicates for each of 100,000 tree replicates. Phylogenetic signal is measured using five signal detection methods. Data are shown from the analysis which uses the same total number of species as all angiosperms and C_3 angiosperms in the rubisco kinetic dataset, respectively.

Number of species	*Cmean*		*I*		*K*		*K**		*Lambda*		Consensus
	Stat	***% Sig***	***Stat***	***% Sig***	***Stat***	***% Sig***	***Stat***	***% Sig***	***Stat***	***% Sig***	***% Sig***
All Angiosperms	0.681 ± 0	100.000	0.121 ± 0	100.000	1.000 ± 0	99.882	1.000 ± 0	100.000	1.012 ± 0	100.000	100.000
C_3 Angiosperms	0.652 ±0	99.999	0.128 ±0	99.997	1.000 ±0	99.717	1.000 ± 0	100.000	1.009 ±0	99.991	99.999

Appendix A, Table S7. The phylogenetic signal strength and associated significance level in rubisco kinetic traits when omitting species in the *Oryza*, *Aegilops* and *Triticum* clades. Phylogenetic signal is measured using five signal detection methods. Data are shown from the analysis of all angiosperms and C_3 angiosperms, respectively.

NOTE. Statistics are rounded to three decimal places and significance values are represented as α levels, where α = 0.001 if P < 0.001, α = 0.01 if 0.001 < P < 0.01, α = 0.05 if 0.01 < P < 0.05, and α = ns if P > 0.05.

Data	*Kinetic Trait*	*Cmean*		*I*		*K*		*K**		*Lambda*	
		Stat	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***
All Angiosperms	$S_{C/O}$	0.527	0.001	0.477	0.001	0.002	0.05	0.002	0.05	0.881	0.001
	k_{catC}	0.320	0.001	0.241	0.001	0.003	0.001	0.004	0.001	0.938	0.01
	K_C	0.285	0.001	0.284	0.001	0.003	0.01	0.004	0.01	0.659	0.001
	K_C^{air}	0.234	0.001	0.268	0.001	0.001	ns	0.001	ns	0.263	ns
	K_O	0.045	ns	- 0.056	ns	0	ns	0	ns	0	ns
C_3 Angiosperms	$S_{C/O}$	0.567	0.001	0.532	0.001	0.002	ns	0.002	ns	0.816	0.001
	k_{catC}	0.353	0.001	0.259	0.001	0.002	ns	0.003	ns	0.762	0.01
	K_C	0.454	0.001	0.402	0.001	0.004	0.05	0.005	0.05	0.946	0.001
	K_C^{air}	0.401	0.001	0.370	0.001	0.003	ns	0.003	ns	0.940	0.001
	K_O	0.346	0.001	0.241	0.001	0.008	0.05	0.009	0.05	0.804	0.001

Appendix A, Table S8. The phylogenetic signal strength and associated significance level in rubisco kinetic traits when assessed using phylogenetic trees inferred from the coding sequence of *matk*, *ndhF*, or *atpB* genes. Phylogenetic signal is measured using five signal detection methods. Data are shown from the analysis of all angiosperms and C_3 angiosperms, respectively.

NOTE. Statistics are rounded to three decimal places and significance values are represented as α levels, where α = 0.001 if P < 0.001, α = 0.01 if 0.001 < P < 0.01, α = 0.05 if 0.01 < P < 0.05, and α = ns if P > 0.05.

Data	Kinetic	Cmean		I		K		K*		Lambda	
	Trait	Stat	α	Stat	α	Stat	α	Stat	α	Stat	α
All	$S_{C/O}$	0.327	0.001	0.267	0.001	0.001	0.01	0.001	0.05	0.855	0.001
Angiosperms	k_{catC}	0.335	0.001	0.21	0.01	0.001	0.01	0.001	0.001	0.845	0.05
matK	K_C	0.097	ns	0.106	ns	0	ns	0	ns	0	ns
(*n* = 85)	K_C^{air}	0.095	ns	0.122	0.05	0	ns	0	ns	0	ns
	K_O	-		-							
		0.058	ns	0.106	ns	0	ns	0	ns	0	ns
All	$S_{C/O}$	0.389	0.001	0.313	0.001	0	0.05	0	ns	0.836	0.001
Angiosperms	k_{catC}	0.429	0.001	0.196	0.01	0.001	0.01	0.001	0.01	0.986	0.001
ndhF	K_C	0.131	0.05	0.052	ns	0	ns	0	ns	0.01	ns
(*n* = 76)	K_C^{air}	0.171	0.05	0.095	ns	0	ns	0	ns	0.001	ns
	K_O	-		-							
		0.116	ns	0.142	ns	0	ns	0	ns	0	ns
All	$S_{C/O}$	0.273	0.001	0.141	0.05	0	ns	0	ns	0.866	0.001
Angiosperms	k_{catC}	0.365	0.001	0.225	0.01	0.001	0.001	0.002	0.001	0.975	0.001
atpB	K_C	0.274	0.001	0.205	0.01	0	0.01	0.001	0.01	0.894	0.05
(*n* = 63)	K_C^{air}	0.205	0.01	0.175	0.05	0.001	0.001	0.001	0.001	0.796	ns
	K_O	0.130	0.05	0.009	ns	0	ns	0	ns	0.122	ns
C_3	$S_{C/O}$	0.308	0.001	0.262	0.01	0.001	0.05	0.001	0.05	0.809	0.001
Angiosperms	k_{catC}	0.396	0.001	0.283	0.001	0.001	0.01	0.001	0.01	0.884	0.001
matK	K_C	0.408	0.001	0.29	0.001	0	0.05	0	0.05	0.865	0.001
(*n* = 69)	K_C^{air}	0.355	0.001	0.257	0.01	0	0.05	0.001	0.05	0.903	0.001
	K_O	0.186	0.05	0.044	ns	0	ns	0	ns	0.376	ns
C_3	$S_{C/O}$	0.111	ns	0.173	0.05	0	ns	0	ns	0.467	0.01
Angiosperms	k_{catC}	0.516	0.001	0.277	0.01	0.001	0.05	0.001	0.05	0.961	0.001
ndhF	K_C	0.419	0.001	0.207	0.01	0	ns	0	ns	0.875	0.001
(*n* = 51)	K_C^{air}	0.437	0.001	0.241	0.01	0	ns	0	ns	0.899	0.001
	K_O	0.048	ns	0.076	ns	0	ns	0	ns	0	ns
C_3	$S_{C/O}$	0.145	0.05	0.051	ns	0	ns	0	ns	0.625	0.001
Angiosperms	k_{catC}	0.51	0.001	0.327	0.001	0.001	0.001	0.001	0.001	0.966	0.001
atpB	K_C	0.364	0.001	0.249	0.01	0	0.01	0.001	0.01	0.911	0.001
(*n* = 52)	K_C^{air}	0.34	0.001	0.253	0.01	0	0.01	0.001	0.001	0.949	0.001
	K_O	0.1	ns	0.001	ns	0	ns	0	ns	0.291	ns

Appendix A, Table S9. The phylogenetic signal strength and associated significance level in rubisco kinetic traits when considering only the subset of rubisco characterized in Orr *et al.*, (2016). Phylogenetic signal is measured using five signal detection methods. Data are shown from the analysis of all angiosperms and C_3 angiosperms, respectively.

NOTE. Statistics are rounded to three decimal places and significance values are represented as α levels, where α = 0.001 if P < 0.001, α = 0.01 if 0.001 < P < 0.01, α = 0.05 if 0.01 < P < 0.05, and α = ns if P > 0.05.

Data	*Kinetic Trait*	*Cmean*		*I*		*K*		*K**		*Lambda*	
		Stat	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***
All Angiosperms (*n* = 69)	$S_{C/O}$	0.183	0.05	0.114	0.05	0.001	ns	0.001	ns	0.274	0.05
	k_{catC}	0.345	0.001	0.241	0.001	0.002	ns	0.003	ns	0.906	0.001
	K_C	0.444	0.001	0.249	0.001	0.002	ns	0.003	ns	0.94	0.001
	K_C^{air}	0.384	0.001	0.21	0.001	0.001	ns	0.002	ns	0.883	0.001
	K_O	0.247	0.01	0.111	0.05	0.009	0.05	0.01	0.05	0.543	0.05
C_3 Angiosperms (*n* = 64)	$S_{C/O}$	0.284	0.001	0.186	0.01	0.001	ns	0.001	ns	0.37	0.001
	k_{catC}	0.403	0.001	0.282	0.001	0.002	ns	0.003	ns	0.874	0.001
	K_C	0.339	0.001	0.278	0.001	0.002	ns	0.002	ns	0.928	0.001
	K_C^{air}	0.293	0.001	0.223	0.001	0.001	ns	0.001	ns	0.819	0.01
	K_O	0.183	0.05	0.113	0.05	0.006	0.05	0.008	0.05	0.464	ns

Appendix A, Table S10. The phylogenetic signal strength and associated significance level in rubisco kinetic traits after a reduction of 13 mol mol^{-1} was applied to $S_{C/O}$ values in *Limonium* species. Phylogenetic signal is measured using five signal detection methods. Data are shown from the analysis of all angiosperms and C_3 angiosperms, respectively.

NOTE. Statistics are rounded to three decimal places and significance values are represented as α levels, where α = 0.001 if P < 0.001, α = 0.01 if 0.001 < P < 0.01, α = 0.05 if 0.01 < P < 0.05, and α = ns if P > 0.05.

Data	Kinetic Trait	Cmean		I		K		K*		Lambda	
		Stat	α	Stat	α	Stat	α	Stat	α	Stat	α
All Angiosperms	$S_{C/O}$	0.396	0.001	0.298	0.001	0.001	ns	0.001	ns	0.818	0.001
	k_{catC}	0.350	0.001	0.224	0.001	0.003	0.001	0.003	0.001	0.968	0.001
	K_C	0.282	0.001	0.248	0.001	0.001	0.01	0.002	0.01	0.902	0.001
	K_C^{air}	0.234	0.001	0.242	0.001	0.001	0.05	0.001	0.05	0.392	ns
	K_O	0.032	ns	-0.07	ns	0	ns	0	ns	0	ns
C_3 Angiosperms	$S_{C/O}$	0.363	0.001	0.255	0.001	0	ns	0	ns	0.678	0.001
	k_{catC}	0.387	0.001	0.234	0.001	0.002	0.01	0.002	0.01	0.913	0.001
	K_C	0.449	0.001	0.341	0.001	0.001	0.05	0.001	0.05	0.948	0.001
	K_C^{air}	0.398	0.001	0.317	0.001	0.001	0.05	0.002	0.01	0.947	0.001
	K_O	0.279	0.001	0.167	0.01	0	ns	0	ns	0.743	0.001

Appendix A, Table S11. The phylogenetic signal strength and associated significance level in rubisco kinetic traits when omitting species in *Limonium* clade. Phylogenetic signal is measured using five signal detection methods. Data are shown from the analysis of all angiosperms and C_3 angiosperms, respectively.

NOTE. Statistics are rounded to three decimal places and significance values are represented as α levels, where α = 0.001 if P < 0.001, α = 0.01 if 0.001 < P < 0.01, α = 0.05 if 0.01 < P < 0.05, and α = ns if P > 0.05.

Data	Kinetic Trait	Cmean		I		K		K*		Lambda	
		Stat	α	Stat	α	Stat	α	Stat	α	Stat	α
All Angiosperms	$S_{C/O}$	0.294	0.001	0.187	0.001	0	ns	0	ns	0.753	0.001
	k_{catC}	0.339	0.001	0.205	0.001	0.003	0.001	0.003	0.001	0.97	0.001
	K_C	0.185	0.01	0.164	0.01	0.001	0.01	0.001	0.01	0.825	ns
	K_C^{air}	0.152	0.01	0.172	0.01	0.001	ns	0.001	0.05	0	ns
	K_O	-		-							
		0.002	ns	0.101	ns	0	ns	0	ns	0	ns
C_3 Angiosperms	$S_{C/O}$	0.220	0.01	0.089	ns	0	ns	0	ns	0.367	0.001
	k_{catC}	0.381	0.001	0.215	0.01	0.002	0.01	0.002	0.01	0.924	0.001
	K_C	0.341	0.001	0.211	0.01	0.001	ns	0.001	0.05	0.943	0.001
	K_C^{air}	0.299	0.001	0.203	0.01	0.001	0.05	0.002	0.05	0.944	0.001
	K_O	0.225	0.01	0.109	0.05	0	ns	0	ns	0.726	0.01

A.8 References

Abouheif, E., 1999. A method for testing the assumption of phylogenetic independence in comparative data. Evol Ecol Res 1, 895–909.

Aguiló-Nicolau, P., Galmés, J., Fais, G., Capó-Bauçà, S., Cao, G., Iñiguez, C., 2023. Singular adaptations in the carbon assimilation mechanism of the polyextremophile cyanobacterium Chroococcidiopsis thermalis. Photosynth Res. https://doi.org/10.1007/S11120-023-01008-Y

Aigner, H., Wilson, R.H., Bracher, A., Calisse, L., Bhat, J.Y., Hartl, F.U., Hayer-Hartl, M., 2017. Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2. Science (1979) 358, 1272–1278. https://doi.org/10.1126/science.aap9221

Anselin, L., 1995. Local Indicators of Spatial Association—LISA. Geogr Anal 27, 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

APG, 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society 181, 1–20. https://doi.org/10.1111/boj.12385

APG, 1998. An Ordinal Classification for the Families of Flowering Plants [WWW Document]. Annals of the Missouri Botanical Garden. https://doi.org/10.2307/2992015

Badger, M.R., Andrews, T.J., Whitney, S.M., Ludwig, M., Yellowlees, D.C., Leggat, W., Price, G.D., 1998. The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO2-concentrating mechanisms in algae. Canadian Journal of Botany 76, 1052–1071. https://doi.org/10.1139/b98-074

Balasubramaniam, K.N., Dittmar, K., Berman, C.M., Butovskaya, M., Cooper, M.A., Majolo, B., Ogawa, H., Schino, G., Thierry, B., de Waal, F.B.M., 2012. Hierarchical steepness and phylogenetic models: Phylogenetic signals in Macaca. Anim Behav 83, 1207–1218. https://doi.org/10.1016/j.anbehav.2012.02.012

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D.S., Milo, R., 2011. The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. Biochemistry 50, 4402–4410. https://doi.org/10.1021/bi2002289

Bathellier, C., Tcherkez, G., Lorimer, G.H., Farquhar, G.D., 2018. Rubisco is not really so bad. Plant Cell Environ. https://doi.org/10.1111/pce.13149

Blomberg, S.P., Garland, T., 2002. Tempo and mode in evolution: Phylogenetic inertia, adaptation and comparative methods. J Evol Biol 15, 899–910. https://doi.org/10.1046/j.1420-9101.2002.00472.x

Blomberg, S.P., Garland, T., Ives, A.R., 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. Evolution (N Y) 57, 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x

Bohs, L., Olmstead, R.G., 1997. Phylogenetic relationships in Solanum (Solanaceae) based on ndhF sequences. Syst Bot 22, 5–17. https://doi.org/10.2307/2419674

Bouvier, J.W., Emms, D.M., Kelly, S., 2022. Slow molecular evolution of rubisco limits adaptive improvement of CO2 assimilation. bioRxiv.

Bouvier, J.W., Emms, D.M., Rhodes, T., Bolton, J.S., Brasnett, A., Eddershaw, A., Nielsen, J.R., Unitt, A., Whitney, S.M., Kelly, S., 2021. Rubisco Adaptation Is More Limited by Phylogenetic Constraint Than by Catalytic Trade-off. Mol Biol Evol 38, 2880–2896. https://doi.org/10.1093/molbev/msab079

Bowes, G., Ogren, W.L., Hageman, R.H., 1971. Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase, Biochemical and Biophysical Research Communications. https://doi.org/10.1016/0006-291X(71)90475-X

Capó-Bauçà, S., Iñiguez, C., Aguiló-Nicolau, P., Galmés, J., 2022. Correlative adaptation between Rubisco and CO2-concentrating mechanisms in seagrasses. Nat Plants 8, 706–716. https://doi.org/10.1038/s41477-022-01171-5

Carmo-Silva, E., Scales, J.C., Madgwick, P.J., Parry, M.A.J., 2015. Optimizing Rubisco and its regulation for greater resource use efficiency. Plant Cell Environ. https://doi.org/10.1111/pce.12425

Chollet, R., 1977. The biochemistry of photorespiration, Trends in Biochemical Sciences. https://doi.org/10.1016/0968-0004(77)90364-4

Cleton-Jansen, A.M., Dekker, S., van de Putte, P., Goosen, N., 1991. A single amino acid substitution changes the substrate specificity of quinoprotein glucose dehydrogenase in Gluconobacter oxydans. MGG Molecular & General Genetics. https://doi.org/10.1007/BF00272157

Cummins, P.L., 2021. The Coevolution of RuBisCO, Photorespiration, and Carbon Concentrating Mechanisms in Higher Plants. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.662425

Cummins, P.L., Kannappan, B., Gready, J.E., 2019. Response: Commentary: Directions for Optimization of Photosynthetic Carbon Fixation: RuBisCO's Efficiency May Not Be So Constrained After All. Front Plant Sci 10, 1426. https://doi.org/10.3389/fpls.2019.01426

Cummins, P.L., Kannappan, B., Gready, J.E., 2018. Directions for optimization of photosynthetic carbon fixation: Rubisco's efficiency may not be so constrained after all. Front Plant Sci 9. https://doi.org/10.3389/fpls.2018.00183

Duraõ, P., Aigner, H., Nagy, P., Mueller-Cajar, O., Hartl, F.U., Hayer-Hartl, M., 2015. Opposing effects of folding and assembly chaperones on evolvability of Rubisco. Nat Chem Biol 11, 148–155. https://doi.org/10.1038/nchembio.1715

Espadaler, J., Eswar, N., Querol, E., Avilés, F.X., Sali, A., Marti-Renom, M.A., Oliva, B., 2008. Prediction of enzyme function by combining sequence similarity and protein interactions. BMC Bioinformatics 9. https://doi.org/10.1186/1471-2105-9-249

Felsenstein, J., 1985. Phylogenies and the comparative method., American Naturalist. https://doi.org/10.1086/284325

Ferguson, D.M., 1998. Phylogenetic analysis and relationships in hydrophyllaceae based on ndhF sequence data. Syst Bot 23, 253–268. https://doi.org/10.2307/2419504

Field, C.B., Behrenfeld, M.J., Randerson, J.T., Falkowski, P., 1998. Primary production of the biosphere: Integrating terrestrial and oceanic components. Science (1979) 281, 237–240. https://doi.org/10.1126/science.281.5374.237

Flamholz, A.I., Prywes, N., Moran, U., Davidi, D., Bar-On, Y.M., Oltrogge, L.M., Alves, R., Savage, D., Milo, R., 2019. Revisiting Trade-offs between Rubisco Kinetic Parameters. Biochemistry. https://doi.org/10.1021/acs.biochem.9b00237

Freckleton, R.P., Harvey, P.H., Pagel, M., 2002. Phylogenetic analysis and comparative data: A test and review of evidence. American Naturalist 160, 712–726. https://doi.org/10.1086/343873

Fritz, S.A., Purvis, A., 2010. Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. Conservation Biology 24, 1042–1051. https://doi.org/10.1111/j.1523-1739.2010.01455.x

Galmés, J., Kapralov, M. v., Andralojc, P.J., Conesa, M.À., Keys, A.J., Parry, M.A.J., Flexas, J., 2014. Expanding knowledge of the Rubisco kinetics variability in plant species: Environmental and evolutionary trends. Plant Cell Environ 37, 1989–2001. https://doi.org/10.1111/pce.12335

Gielly, L., Taberlet, P., 1994. The use of chloroplast DNA to resolve plant phylogenies: Noncoding versus rbcL sequences, Molecular Biology and Evolution. https://doi.org/10.1093/oxfordjournals.molbev.a040157

Gittleman, J.L., Kot, M., 1990. Adaptation: Statistics and a Null Model for Estimating Phylogenetic Effects. Syst Biol 39, 227–241. https://doi.org/10.2307/2992183

Grafen, A., 1989. The phylogenetic regression. Philos Trans R Soc Lond B Biol Sci 326, 119–157. https://doi.org/10.1098/rstb.1989.0106

Hansen, T.F., Martins, E.P., 1996. Translating between microevolutionary process and macroevolutionary patterns: The correlation structure of interspecific data. Evolution (N Y) 50, 1404–1417. https://doi.org/10.1111/j.1558-5646.1996.tb03914.x

Hilu, K.W., Liang, H., 1997. The matK gene sequence variation and application in plant systematics. Am J Bot 84, 830–839. https://doi.org/10.2307/2445819

Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. Syst Biol. https://doi.org/10.1093/sysbio/sys062

Iñiguez, C., Capó-Bauçà, S., Niinemets, Ü., Stoll, H., Aguiló-Nicolau, P., Galmés, J., 2020. Evolutionary trends in RuBisCO kinetics and their co-evolution with CO2 concentrating mechanisms. Plant Journal 101, 897–918. https://doi.org/10.1111/tpj.14643

Iñiguez, C., Niinemets, Ü., Mark, K., Galmés, J., 2021. Analyzing the causes of method-to-method variability among Rubisco kinetic traits: From the first to the current measurements. J Exp Bot 72, 7846–7862. https://doi.org/10.1093/jxb/erab356

Johnson, E.T., Ryu, S., Yi, H., Shin, B., Cheong, H., Choi, G., 2001. Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase. Plant Journal. https://doi.org/10.1046/j.1365-313X.2001.00962.x

Kamilar, J.M., Cooper, N., 2013. Phylogenetic signal in primate behaviour, ecology and life history. Philosophical Transactions of the Royal Society B: Biological Sciences 368. https://doi.org/10.1098/rstb.2012.0341

Kane, H.J., Viil, J., Entsch, B., Paul, K., Morrell, M.K., Andrews, T.J., 1994. An improved method for measuring the CO2/O2 specificity of ribulosebisphosphatte carboxylase-oxygenase. Aust J Plant Physiol 21, 449–461. https://doi.org/10.1071/PP9940449

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. https://doi.org/10.1093/molbev/mst010

Kelly, S., 2018. The amount of nitrogen used for photosynthesis modulates molecular evolution in plants. Mol Biol Evol 35, 1616–1625. https://doi.org/10.1093/molbev/msy043

Koch, M., Haubold, B., Mitchell-Olds, T., 2001. Molecular systematics of the brassicaceae: Evidence from coding plastidic matK and nuclear Chs sequences. Am J Bot 88, 534–544. https://doi.org/10.2307/2657117

Kubien, D.S., Whitney, S.M., Moore, P. v., Jesson, L.K., 2008. The biochemistry of Rubisco in Flaveria. J Exp Bot 59, 1767–1777. https://doi.org/10.1093/jxb/erm283

Lindenfors, P., Revell, L.J., Nunn, C.L., 2010. Sexual dimorphism in primate aerobic capacity: A phylogenetic test. J Evol Biol 23, 1183–1194. https://doi.org/10.1111/j.1420-9101.2010.01983.x

Losos, J.B., 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. Ecol Lett. https://doi.org/10.1111/j.1461-0248.2008.01229.x

Lukas, D., Towner, M., Borgerhoff Mulder, M., 2021. The potential to infer the historical pattern of cultural macroevolution. Philosophical Transactions of the Royal Society B: Biological Sciences 376. https://doi.org/10.1098/rstb.2020.0057

Martins, E.P., Hansen, T.F., 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. American Naturalist 149, 646–667. https://doi.org/10.1086/286013

Motani, R., Schmitz, L., 2011. Phylogenetic versus functional signals in the evolution of form-function relationships in terrestrial vision. Evolution (N Y) 65, 2245–2257. https://doi.org/10.1111/j.1558-5646.2011.01271.x

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., Thuiller, W., 2012. How to measure and test phylogenetic signal. Methods Ecol Evol 3, 743–756. https://doi.org/10.1111/j.2041-210X.2012.00196.x

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. https://doi.org/10.1093/molbev/msu300

Occhialini, A., Lin, M.T., Andralojc, P.J., Hanson, M.R., Parry, M.A.J., 2016. Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO2. Plant Journal 85, 148–160. https://doi.org/10.1111/tpj.13098

Orr, D.J., Alcântara, A., Kapralov, M. v., John Andralojc, P., Carmo-Silva, E., Parry, M.A.J., 2016. Surveying rubisco diversity and temperature response to improve crop photosynthetic efficiency. Plant Physiol 172, 707–717. https://doi.org/10.1104/pp.16.00750

Pagel, M., 1999. Inferring the historical patterns of biological evolution. Nature. https://doi.org/10.1038/44766

Pagel, M., 1997. Inferring evolutionary processes from phylogenies. Zool Scr 26, 331–348. https://doi.org/10.1111/j.1463-6409.1997.tb00423.x

Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290. https://doi.org/10.1093/bioinformatics/btg412

Paradis, E., Schliep, K., 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528. https://doi.org/10.1093/bioinformatics/bty633

Parry, M.A.J., Keys, A.J., Gutteridge, S., 1989. Variation in the specificity factor of C3 higher plant rubiscos determined by the total consumption of ribulose-P2. J Exp Bot 40, 317–320. https://doi.org/10.1093/jxb/40.3.317

Petersen, G., Seberg, O., Yde, M., Berthelsen, K., 2006. Phylogenetic relationships of Triticum and Aegilops and evidence for the origin of the A, B, and D genomes of common wheat (Triticum aestivum). Mol Phylogenet Evol 39, 70–82. https://doi.org/10.1016/j.ympev.2006.01.023

Prins, A., Orr, D.J., Andralojc, P.J., Reynolds, M.P., Carmo-Silva, E., Parry, M.A.J., 2016. Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis. J Exp Bot 67, 1827–1838. https://doi.org/10.1093/jxb/erv574

Revell, L.J., 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Revell, L.J., 2009. Size-correction and principal components for interspecific comparative studies. Evolution (N Y) 63, 3258–3268. https://doi.org/10.1111/j.1558-5646.2009.00804.x

Robbins, E.H.J., Kelly, S., 2022. The evolutionary constraints on angiosperm chloroplast adaptation. bioRxiv 2022.07.12.499704.

Rohlf, F.J., 2001. Comparative methods for the analysis of continuous variables: Geometric interpretations, Evolution. https://doi.org/10.1111/j.0014-3820.2001.tb00731.x

Savir, Y., Noor, E., Milo, R., Tlusty, T., 2010. Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. Proc Natl Acad Sci U S A 107, 3475–3480. https://doi.org/10.1073/pnas.0911663107

Savolainen, V., Chase, M.W., Hoot, S.B., Morton, C.M., Soltis, D.E., Bayer, C., Fay, M.F., de Bruijn, A.Y., Sullivan, S., Qiu, Y.L., 2000. Phylogenetics of Flowering Plants Based on Combined Analysis of Plastid atpB and rbcL Gene Sequences. Syst Biol 49, 306–362. https://doi.org/10.1080/10635159950173861

Seward, E.A., Kelly, S., 2018. Selection-driven cost-efficiency optimization of transcripts modulates gene evolutionary rate in bacteria. Genome Biol 19. https://doi.org/10.1186/s13059-018-1480-7

Sharkey, T.D., 2020. Emerging research in plant photosynthesis. Emerg Top Life Sci 4, 137–150. https://doi.org/10.1042/ETLS20200035

Sharwood, R.E., Ghannoum, O., Kapralov, M. v., Gunn, L.H., Whitney, S.M., 2016. Temperature responses of Rubisco from Paniceae grasses provide opportunities for improving C 3 photosynthesis. Nat Plants 2. https://doi.org/10.1038/nplants.2016.186

Studer, R.A., Christin, P.A., Williams, M.A., Orengo, C.A., 2014. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. Proc Natl Acad Sci U S A 111, 2223–2228. https://doi.org/10.1073/pnas.1310811111

Symonds, M.R.E., Blomberg, S.P., 2014. A primer on phylogenetic generalised least squares, Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-43550-2_5

Tabita, F.R., Hanson, T.E., Satagopan, S., Witte, B.H., Kreel, N.E., 2008. Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 2629–2640. https://doi.org/10.1098/rstb.2008.0023

Tcherkez, G., Farquhar, G.D., 2021. Rubisco catalytic adaptation is mostly driven by photosynthetic conditions – Not by phylogenetic constraints. J Plant Physiol 267. https://doi.org/10.1016/j.jplph.2021.153554

Tcherkez, G.G.B., Farquhar, G.D., Andrews, T.J., 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. Proc Natl Acad Sci U S A 103, 7246–7251. https://doi.org/10.1073/pnas.0600605103

van den Boogaart, K.G., Tolosana-Delgado, R., 2008. "compositions": A unified R package to analyze compositional data. Comput Geosci 34, 320–338. https://doi.org/10.1016/j.cageo.2006.11.017

Villar, K. Del, Mitsuzawa, H., Yang, W., Sattler, I., Tamanoi, F., 1997. Amino acid substitutions that convert the protein substrate specificity of farnesyltransferase to that of geranylgeranyltransferase type I. Journal of Biological Chemistry. https://doi.org/10.1074/jbc.272.1.680

von Caemmerer, S., Evans, J.R., Hudson, G.S., Andrews, T.J., 1994. The kinetics of ribulose-1,5-bisphosphate carboxylase/oxygenase in vivo inferred from measurements of photosynthesis in leaves of transgenic tobacco. Planta 195, 88–97. https://doi.org/10.1007/BF00206296

Whitney, S.M., Sharwood, R.E., Orr, D., White, S.J., Alonso, H., Galmés, J., 2011. Isoleucine 309 acts as a C 4 catalytic switch that increases ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) carboxylation rate in flaveria. Proc Natl Acad Sci U S A 108, 14688–14693. https://doi.org/10.1073/PNAS.1109503108/-/DCSUPPLEMENTAL

Wiens, J.J., Kuczynski, C.A., Smith, S.A., Mulcahy, D.G., Sites, J.W., Townsend, T.M., Reeder, T.W., 2008. Branch lengths, support, and congruence: Testing the phylogenomic approach with 20 nuclear loci in snakes. Syst Biol 57, 420–431. https://doi.org/10.1080/10635150802166053

Wilson, R.H., Martin-Avila, E., Conlan, C., Whitney, S.M., 2018. An improved Escherichia coli screen for Rubisco identifies a protein-protein interface that can enhance CO2-fixation kinetics. Journal of Biological Chemistry 293, 18–27. https://doi.org/10.1074/jbc.M117.810861

Wolf, P.G., 1997. Evaluation of atpB nucleotide sequences for phylogenetic studies of ferns and other pteridophytes. Am J Bot 84, 1429–1440. https://doi.org/10.2307/2446141

Wu, E.T.Y., Liu, Y., Jennings, L., Dong, S., Davies, T.J., 2021. Detecting the phylogenetic signal of glacial refugia in a bryodiversity hotspot outside the tropics. Divers Distrib. https://doi.org/10.1111/ddi.13449

Yen, A.C., Olmstead, R.G., 2000. Molecular systematics of Cyperaceae tribe Cariceae based on two chloroplast DNA regions: ndhF and trnL intron-intergenic spacer. Syst Bot 25, 479–494. https://doi.org/10.2307/2666691

Young, J.N., Heureux, A.M.C., Sharwood, R.E., Rickaby, R.E.M., Morel, F.M.M., Whitney, S.M., 2016. Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentrating mechanisms. J Exp Bot 67, 3445–3456. https://doi.org/10.1093/jxb/erw163

Zhou, Y., Whitney, S., 2019. Directed evolution of an improved Rubisco; in vitro analyses to decipher fact from fiction. Int J Mol Sci 20. https://doi.org/10.3390/ijms20205019

Zhu, G., Jensen, R.G., Bohnert, H.J., Wildner, G.F., Schlitter, J., 1998. Dependence of catalysis and CO2/O2 specificity of Rubisco on the carboxy-terminus of the large subunit at different temperatures. Photosynth Res 57, 71–79. https://doi.org/10.1023/A:1006020424581

Appendix B, Supplemental Data and Text

Supporting Chapter 2

Chapter 2, Contributors:
Bouvier, J.W., Emms, D.M., Rhodes, T., Bolton, J.S., Brasnett, A., Eddershaw, A., Nielsen, J.R., Unitt, A., Whitney, S.M. and Kelly, S.

Author contributions to data Chapter 2:
S.K. conceived the study. J.W.B. conducted the analysis. J.W.B., T.R., S.M.W., D.M.E., and S.K. analyzed the data. J.W.B. and S.K. wrote the manuscript. J.R.N., J.S.B., A.E., A.B., and A.U. contributed to discussions and methodological development. All authors edited the manuscript.

This thesis chapter has been published at the following location:
Bouvier, J.W., et al., 2021. Molecular Biology and Evolution, 38(7), 2880-2896

All supplemental data files are provided in the online version of this manuscript.

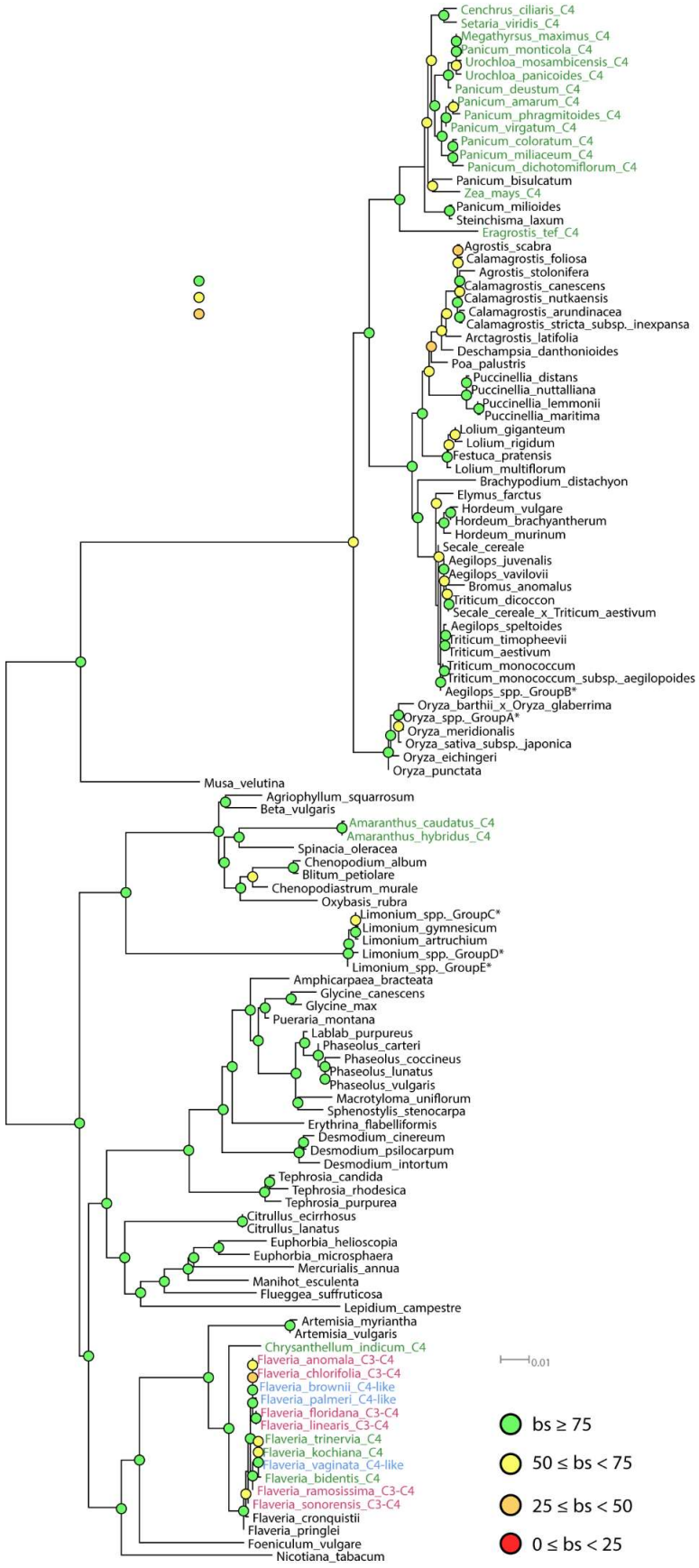
B.1 Supplemental Results

B.1.1 The phylogenetic signal in rubisco kinetic traits is not caused by overfitting due to use of rbcL for phylogenetic tree inference

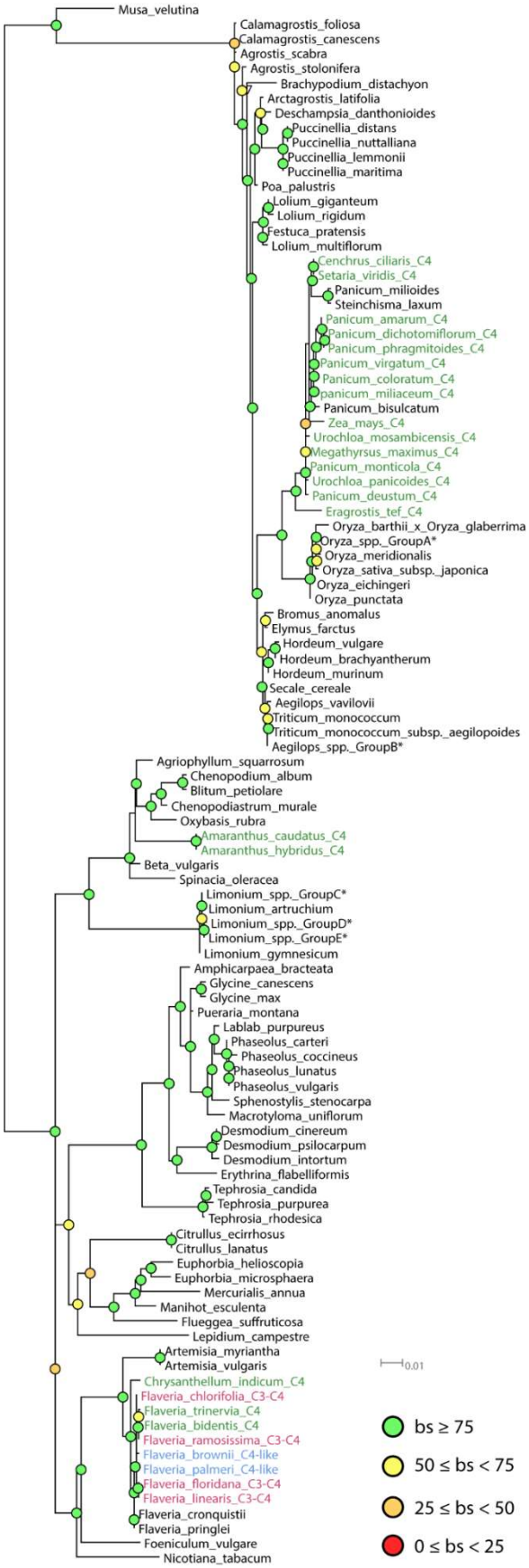
Although a phylogenetic tree inferred from the nucleotide sequence of the *rbcL* gene is widely accepted as a good proxy for the true phylogenetic relationships between plant species (Gielly and Taberlet, 1994), and is used in many large scale studies of plant phylogeny (APG, 2016), there is a possibility that the use of this gene when assessing for a phylogenetic signal in rubisco kinetic traits could have resulted in overfitting. Specifically, amino acid substitutions that contribute to variation in measured rubisco kinetics would also contribute to the topology of the phylogenetic tree. Thus, to test for the presence of this potential overfitting and prevent such amino acid substitutions influencing

the phylogenetic relationship between species, a phylogenetic tree was inferred for all species from a nucleotide sequence alignment of the *rbcL* gene using only ubiquitously conserved amino acid positions (i.e. sites where the amino acid was constant across all species, but the coding sequence exhibited synonymous sequence changes). The use of this phylogenetic tree containing no amino acid sequence variation (Appendix B, Figure S2) did not affect the conclusions of phylogenetic signal methods in any kinetic trait assessed. Here, an overall significant phylogenetic signal was found in all kinetic traits except K_O (Appendix B, table S1), analogous to results previously described based on the tree inferred from the complete *rbcL* coding sequence (**t**able 2.1). Therefore, measurements of the phylogenetic signal are not an artefact of using the *rbcL* gene to infer the phylogenetic tree.

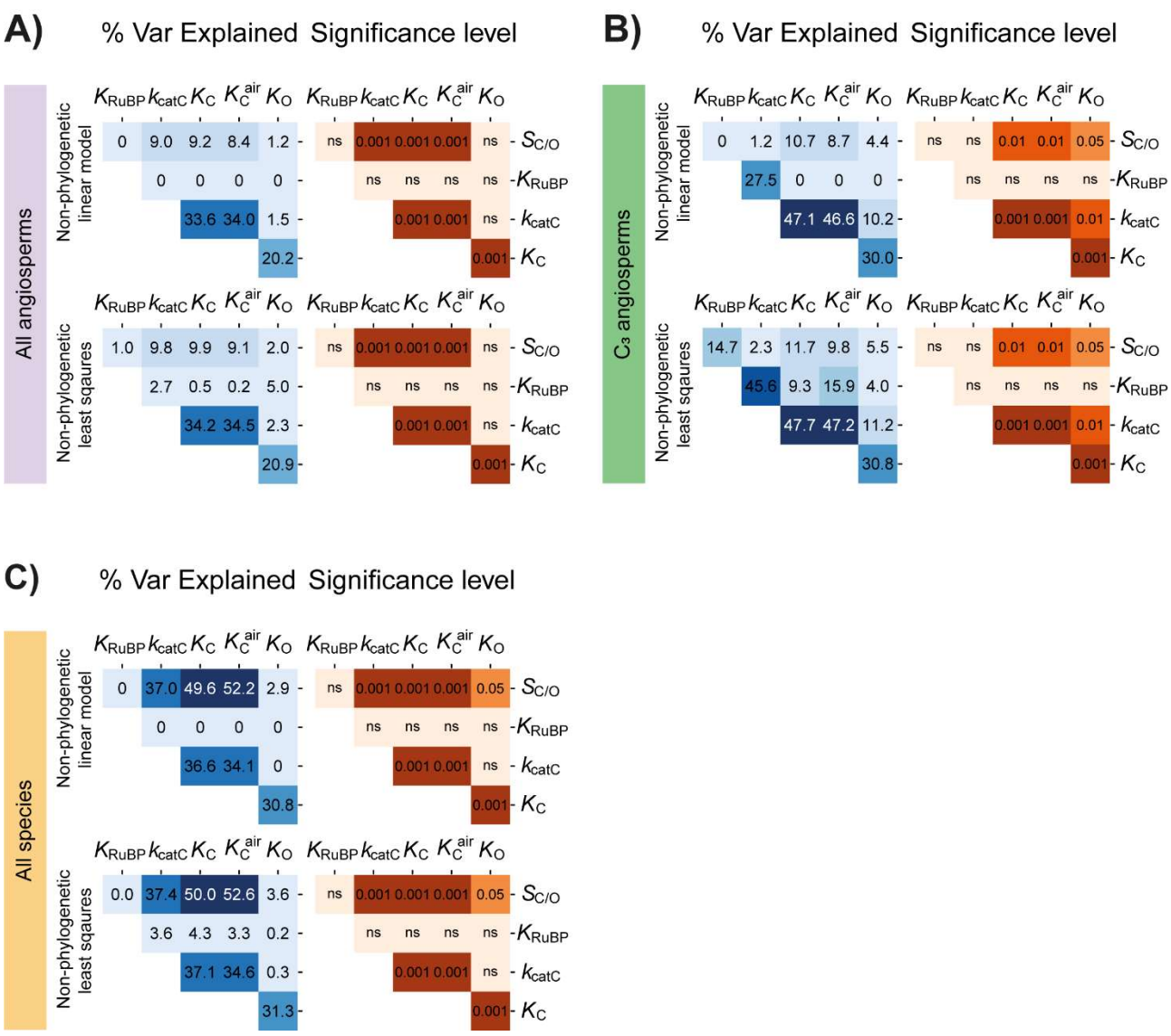
B.2 Supplemental Figures



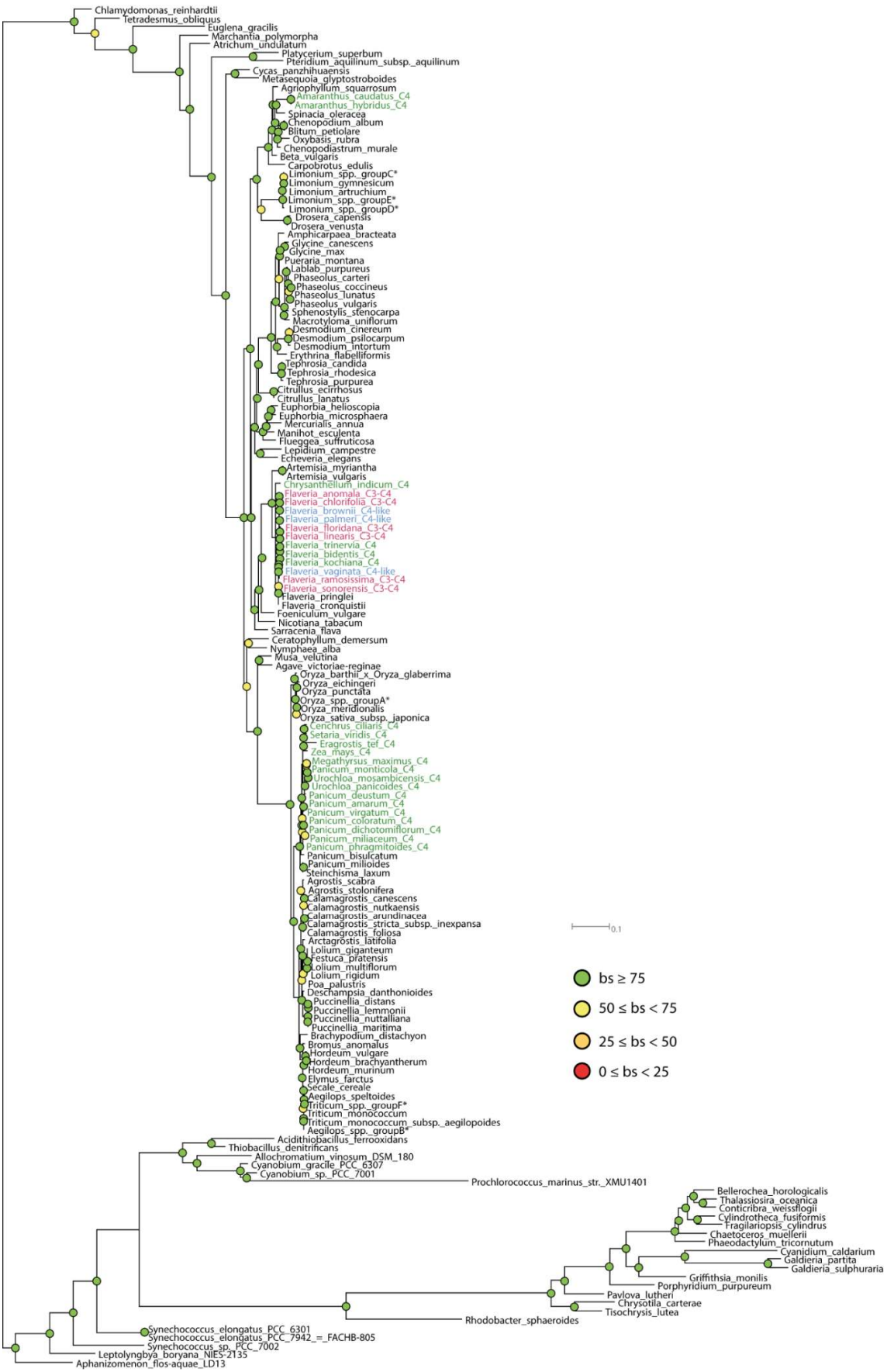
Appendix B, Figure S1. Consensus phylogenetic tree of the angiosperm species in this study inferred from the nucleotide sequence of the *rbcL* gene of the large subunit of the rubisco enzyme. Bootstrap support (bs) values are indicated at internal nodes with scale indicated next to the tree. Where known, photosynthetic types other than the ancestral C_3 pathway have been annotated on to species labels which are colour coded for visualisation, and include C_3-C_4 intermediates (red), C_4-like (light green) and C_4 plants (green). Species which shared identical gene sequences have been condensed into single nodes on the tree to avoid terminal zero-length branches and are signified by asterisks. Of these, *Oryza* spp. Group A include *O. longistaminata*, *O. glaberrima*, *O. sativa* f. *spontanea*, *O. sativa* subpsp. *indica* and *O. glumipatula*; *Aegilops* spp. Group B include *A. triuncialis*, *A. uniaristata*, *A. tauschii*, *A. comosa*, *A. biuncialis*, *A. cylindrica*; *Limonium* spp. Group C include *L. antonii-llorensii*, *L. gibertii*, *L. biflorum*; *Limonium* spp. Group D include *L. echioides*, *L. barceloi*, *L. balearicum*, *L. companyonis*; and *Limonium* spp. Group E include *L. ejulabilis*, *L. retusum*, *L. leonardi-llorensii*, *L. magallufianum*, *L. grosii*. **[*End of Figure Legend*].**


Musa_velutina
Calamagrostis_foliosa
Calamagrostis_canescens
Agrostis_scabra
Agrostis_stolonifera
Brachypodium_distachyon
Arctagrostis_latifolia
Deschampsia_danthonioides
Puccinellia_distans
Puccinellia_nuttalliana
Puccinellia_lemmonii
Puccinellia_maritima
Poa_palustris
Lolium_giganteum
Lolium_rigidum
Festuca_pratensis
Lolium_multiflorum
Cenchrus_ciliaris_C4
Setaria_viridis_C4
Panicum_milioides
Steinchisma_laxum
Panicum_amarum_C4
Panicum_dichotomiflorum_C4
Panicum_phragmitoides_C4
Panicum_virgatum_C4
Panicum_coloratum_C4
panicum_miliaceum_C4
Panicum_bisulcatum
Zea_mays_C4
Urochloa_mosambicensis_C4
Megathyrsus_maximus_C4
Panicum_monticola_C4
Urochloa_panicoides_C4
Panicum_deustum_C4
Eragrostis_tef_C4
Oryza_barthii_x_Oryza_glaberrima
Oryza_spp._GroupA*
Oryza_meridionalis
Oryza_sativa_subsp._japonica
Oryza_eichingeri
Oryza_punctata
Bromus_anomalus
Elymus_farctus
Hordeum_vulgare
Hordeum_brachyantherum
Hordeum_murinum
Secale_cereale
Aegilops_vavilovii
Triticum_monococcum
Triticum_monococcum_subsp._aegilopoides
Aegilops_spp._GroupB*
Agriophyllum_squarrosum
Chenopodium_album
Blitum_petiolare
Chenopodiastrum_murale
Oxybasis_rubra
Amaranthus_caudatus_C4
Amaranthus_hybridus_C4
Beta_vulgaris
Spinacia_oleracea
Limonium_spp._GroupC*
Limonium_artruchium
Limonium_spp._GroupD*
Limonium_spp._GroupE*
Limonium_gymnesicum
Amphicarpaea_bracteata
Glycine_canescens
Glycine_max
Pueraria_montana
Lablab_purpureus
Phaseolus_carteri
Phaseolus_coccineus
Phaseolus_lunatus
Phaseolus_vulgaris
Sphenostylis_stenocarpa
Macrotyloma_uniflorum
Desmodium_cinereum
Desmodium_psilocarpum
Desmodium_intortum
Erythrina_flabelliformis
Tephrosia_candida
Tephrosia_purpurea
Tephrosia_rhodesica
Citrullus_ecirrhosus
Citrullus_lanatus
Euphorbia_helioscopia
Euphorbia_microsphaera
Mercurialis_annua
Manihot_esculenta
Flueggea_suffruticosa
Lepidium_campestre
Artemisia_myriantha
Artemisia_vulgaris
Chrysanthellum_indicum_C4
Flaveria_chlorifolia_C3-C4
Flaveria_trinervia_C4
Flaveria_bidentis_C4
Flaveria_ramosissima_C3-C4
Flaveria_brownii_C4-like
Flaveria_palmeri_C4-like
Flaveria_floridana_C3-C4
Flaveria_linearis_C3-C4
Flaveria_cronquistii
Flaveria_pringlei
Foeniculum_vulgare
Nicotiana_tabacum
0.01
bs ≥ 75
50 ≤ bs < 75
25 ≤ bs < 50
0 ≤ bs < 25

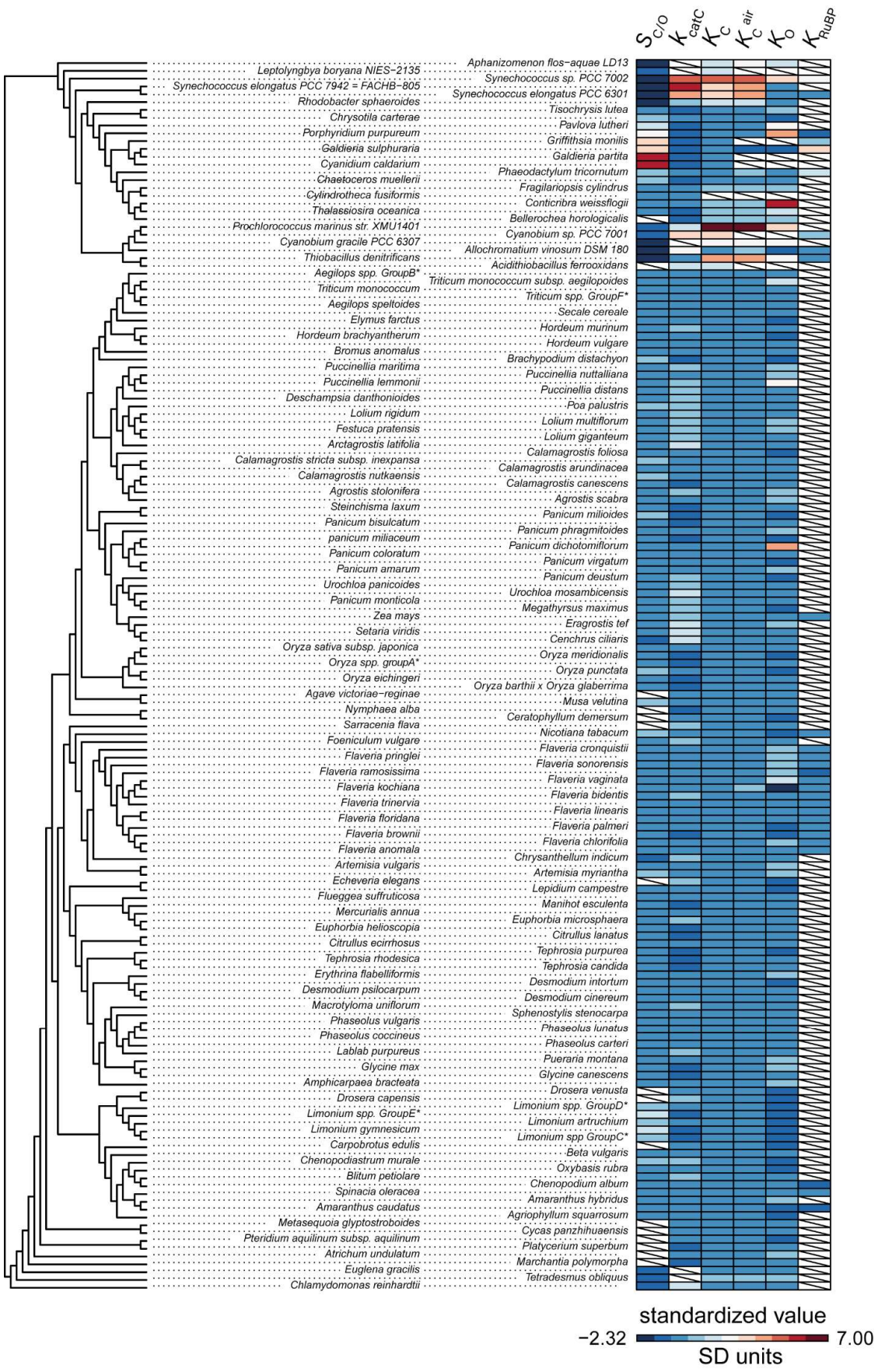
Appendix B, Figure S2. Consensus phylogenetic tree of the angiosperm species in this study inferred from the nucleotide sequence of the *rbcL* gene of the large subunit of the rubisco enzyme when including only sites which encode ubiquitously conserved amino acid positions. The confidence of branch points on the tree are depicted as nodes which are colour coded by bootstrap support (bs) values. Where known, photosynthetic types other than the ancestral C_3 pathway have been annotated on to species labels which are colour coded for visualisation, and include C_3-C_4 intermediates (red), C_4-like (blue) and C_4 plants (green). Species which shared identical sequences when considering the entire sequence length have been condensed into single nodes on the tree to avoid terminal zero-length branches and are signified by asterisks (see Appendix B, Figure S1 for further details). In addition to these, a number of terminal zero length branches were present between nodes on the tree as an artefact of removing non-synonymous positions in the *rbcL* sequence alignment. All but one of the species within these respective groups of plants containing zero length branches were removed and are thus not present in this tree, including *Aegilops juvenalis*, *Aegilops speltoides*, *Calamagrostis arundinacea*, *Calamagrostis stricta subsp. inexpansa*, *Calamagrostis nutkaensis*, *Flaveria anomala*, *Flaveria kochiana*, *Flaveria sonorensis*, *Flaveria vaginata*, *Triticum dicoccon*, *Triticum timopheevii*, *Secale cereale x Triticum aestivum* and *Triticum aestivum.* **[*End of Figure Legend*].**



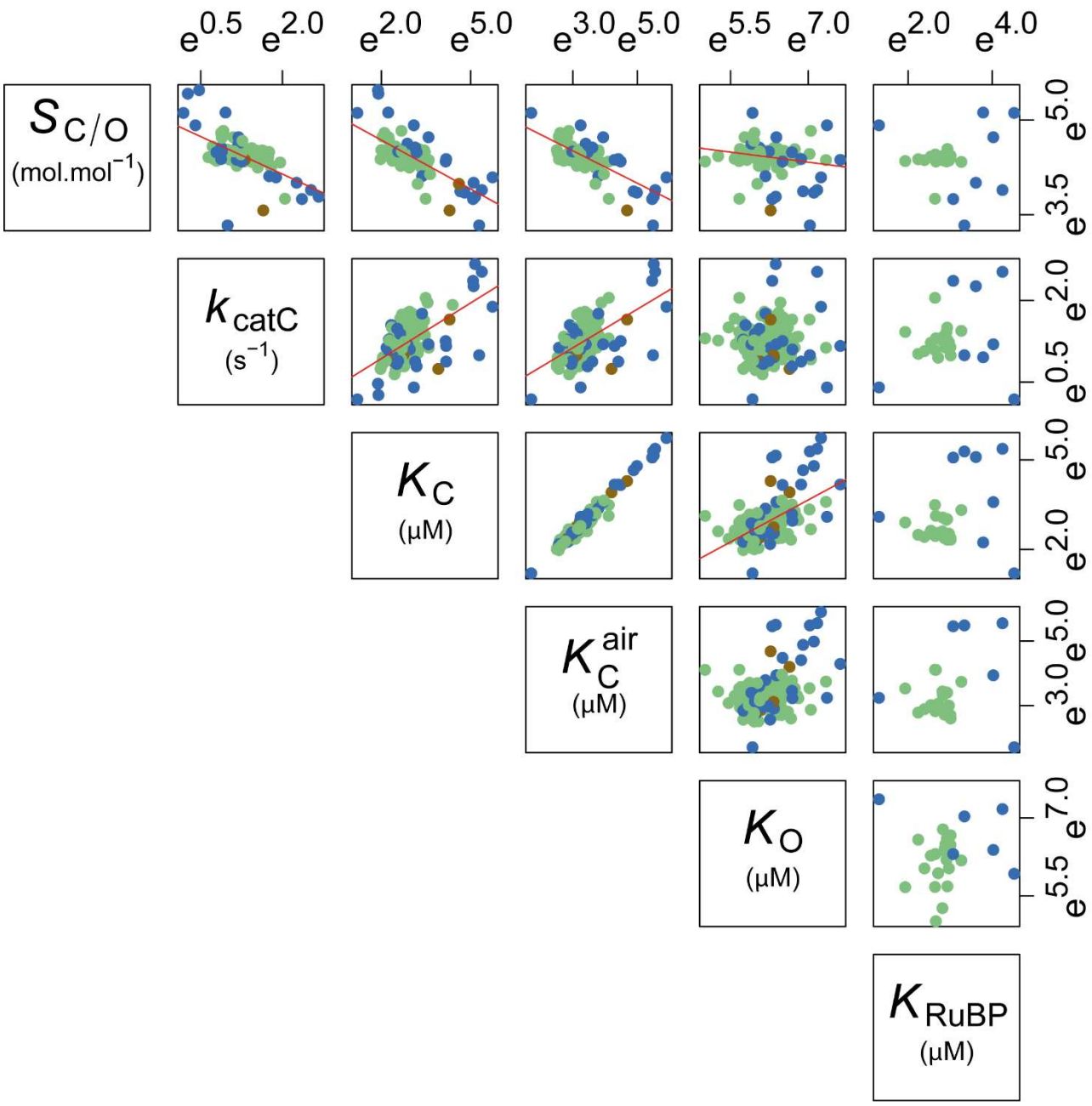
Appendix B, Figure S3. A) Pairwise correlation coefficients (percent variance explained) and associated p-values between different rubisco kinetic traits assessed across the complete set of angiosperms using either non-phylogenetic linear models or least squares regression models. Significance values are represented as α levels, where; α = 0.001 if $p < 0.001$, α = 0.01 if $0.001 < p < 0.01$, α = 0.05 if $0.01 < p < 0.05$, and α = ns if $p > 0.05$. **B)** as in A but for C_3 angiosperms only. **C)** as in A and B but for the complete set of photosynthetic organisms.


Chlamydomonas_reinhardtii
Tetradesmus_obliquus
Euglena_gracilis
Marchantia_polymorpha
Atrichum_undulatum
Platycerium_superbum
Pteridium_aquilinum_subsp._aquilinum
Cycas_panzhihuaensis
Metasequoia_glyptostroboides
Agriophyllum_squarrosum
Amaranthus_caudatus_C4
Amaranthus_hybridus_C4
Spinacia_oleracea
Chenopodium_album
Blitum_petiolare
Oxybasis_rubra
Chenopodiastrum_murale
Beta_vulgaris
Carpobrotus_edulis
Limonium_spp._groupC*
Limonium_gymnesicum
Limonium_artruchium
Limonium_spp._groupE*
Limonium_spp._groupD*
Drosera_capensis
Drosera_venusta
Amphicarpaea_bracteata
Glycine_canescens
Glycine_max
Pueraria_montana
Lablab_purpureus
Phaseolus_carteri
Phaseolus_coccineus
Phaseolus_lunatus
Phaseolus_vulgaris
Sphenostylis_stenocarpa
Macrotyloma_uniflorum
Desmodium_cinereum
Desmodium_psilocarpum
Desmodium_intortum
Erythrina_flabelliformis
Tephrosia_candida
Tephrosia_rhodesica
Tephrosia_purpurea
Citrullus_ecirrhosus
Citrullus_lanatus
Euphorbia_helioscopia
Euphorbia_microsphaera
Mercurialis_annua
Manihot_esculenta
Flueggea_suffruticosa
Lepidium_campestre
Echeveria_elegans
Artemisia_myriantha
Artemisia_vulgaris
Chrysanthellum_indicum_C4
Flaveria_anomala_C3-C4
Flaveria_chlorifolia_C3-C4
Flaveria_brownii_C4-like
Flaveria_palmeri_C4-like
Flaveria_floridana_C3-C4
Flaveria_linearis_C3-C4
Flaveria_trinervia_C4
Flaveria_bidentis_C4
Flaveria_kochiana_C4
Flaveria_vaginata_C4-like
Flaveria_ramosissima_C3-C4
Flaveria_sonorensis_C3-C4
Flaveria_pringlei
Flaveria_cronquistii
Foeniculum_vulgare
Nicotiana_tabacum
Sarracenia_flava
Ceratophyllum_demersum
Nymphaea_alba
Musa_velutina
Agave_victoriae-reginae
Oryza_barthii_x_Oryza_glaberrima
Oryza_eichingeri
Oryza_punctata
Oryza_spp._groupA*
Oryza_meridionalis
Oryza_sativa_subsp._japonica
Cenchrus_ciliaris_C4
Setaria_viridis_C4
Eragrostis_tef_C4
Zea_mays_C4
Megathyrsus_maximus_C4
Panicum_monticola_C4
Urochloa_mosambicensis_C4
Urochloa_panicoides_C4
Panicum_deustum_C4
Panicum_amarum_C4
Panicum_virgatum_C4
Panicum_coloratum_C4
Panicum_dichotomiflorum_C4
Panicum_miliaceum_C4
Panicum_phragmitoides_C4
Panicum_bisulcatum
Panicum_milioides
Steinchisma_laxum
Agrostis_scabra
Agrostis_stolonifera
Calamagrostis_canescens
Calamagrostis_nutkaensis
Calamagrostis_arundinacea
Calamagrostis_stricta_subsp._inexpansa
Calamagrostis_foliosa
Arctagrostis_latifolia
Lolium_giganteum
Festuca_pratensis
Lolium_multiflorum
Lolium_rigidum
Poa_palustris
Deschampsia_danthonioides
Puccinellia_distans
Puccinellia_lemmonii
Puccinellia_nuttalliana
Puccinellia_maritima
Brachypodium_distachyon
Bromus_anomalus
Hordeum_vulgare
Hordeum_brachyantherum
Hordeum_murinum
Elymus_farctus
Secale_cereale
Aegilops_speltoides
Triticum_spp._groupF*
Triticum_monococcum
Triticum_monococcum_subsp._aegilopoides
Aegilops_spp._groupB*
Acidithiobacillus_ferrooxidans
Thiobacillus_denitrificans
Allochromatium_vinosum_DSM_180
Cyanobium_gracile_PCC_6307
Cyanobium_sp._PCC_7001
Prochlorococcus_marinus_str._XMU1401
Bellerochea_horologicalis
Thalassiosira_oceanica
Conticribra_weissflogii
Cylindrotheca_fusiformis
Fragilariopsis_cylindrus
Chaetoceros_muellerii
Phaeodactylum_tricornutum
Cyanidium_caldarium
Galdieria_partita
Galdieria_sulphuraria
Griffithsia_monilis
Porphyridium_purpureum
Pavlova_lutheri
Chrysotila_carterae
Tisochrysis_lutea
Rhodobacter_sphaeroides
Synechococcus_elongatus_PCC_6301
Synechococcus_elongatus_PCC_7942_=_FACHB-805
Synechococcus_sp._PCC_7002
Leptolyngbya_boryana_NIES-2135
Aphanizomenon_flos-aquae_LD13
0.1
bs ≥ 75
50 ≤ bs < 75
25 ≤ bs < 50
0 ≤ bs < 25

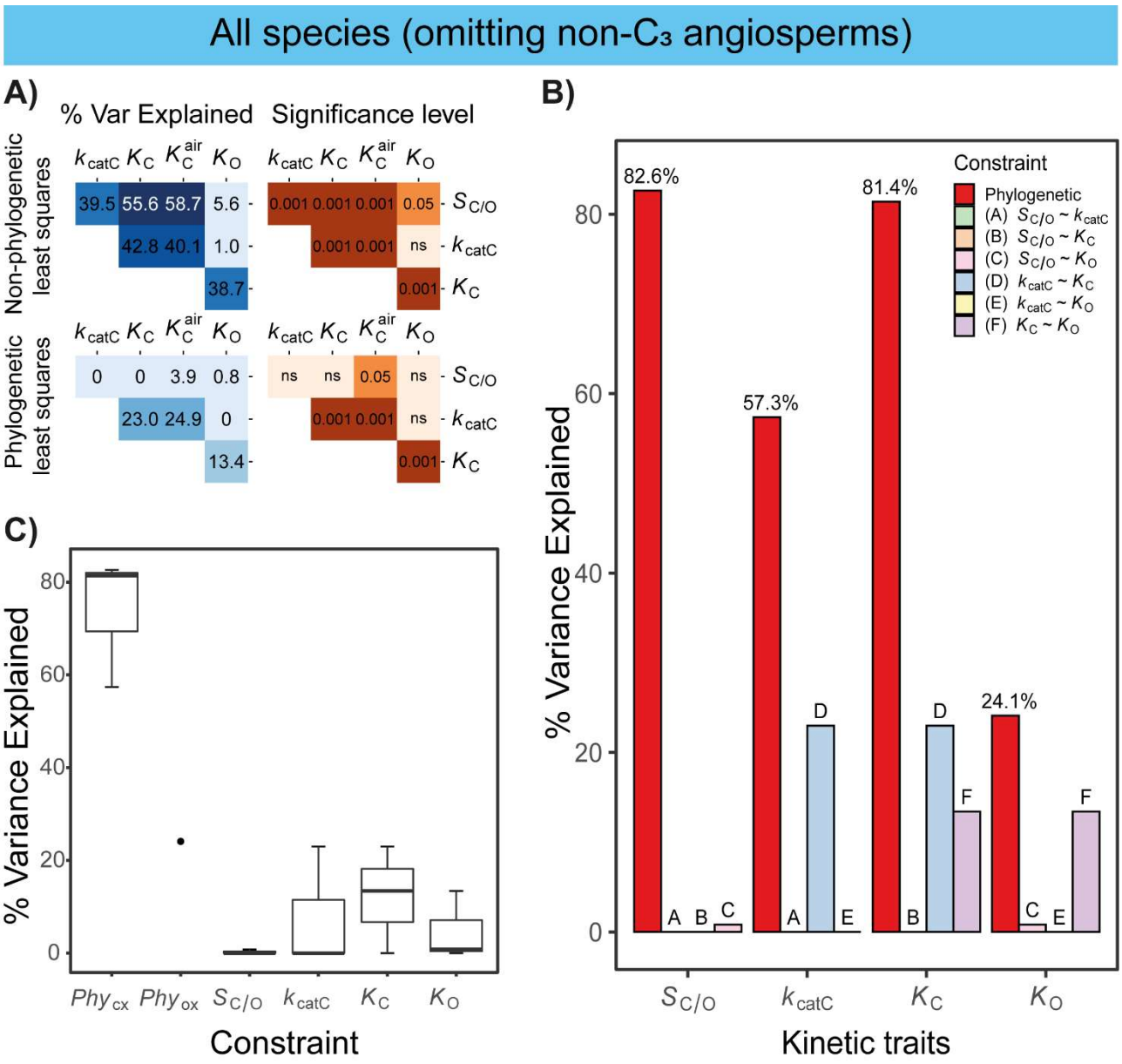
Appendix B, Figure S4. Consensus phylogenetic tree of all photosynthetic organisms in this study inferred from the nucleotide sequence of the *rbcL* gene of the large subunit of the rubisco enzyme. The confidence of branch points on the tree are depicted as nodes which are colour coded by bootstrap support (bs) values. Where known, photosynthetic types other than the ancestral C_3 pathway have been annotated on to species labels which are colour coded for visualisation, and include C_3-C_4 intermediates (red), C_4-like (blue) and C_4 plants (green). Species which shared identical sequences when considering the entire sequence length have been condensed into single nodes on the tree to avoid terminal zero-length branches and are signified by asterisks (see Appendix B, Figure S1 for further details). In addition to those species mentioned in Appendix B, Figure S1, an additional number of species exhibited terminal zero length branches relative to other nodes in this tree due to more stringent trimming of non-aligned positions at the terminal ends of the sequence alignment of all photosynthetic organisms. Of these additional species which exhibited terminal zero length branches, *Aegilops juvenalis* and *Aegilops vavilovii* were consensed into the node *Aegilops* spp. Group B, and *Triticum dicoccon, Secale cereale x Triticum aestivum*, *Triticum aestivum* and *Triticum timopheevii* were condensed into the node *Triticum* spp. Group F. **[*End of Figure Legend*].**



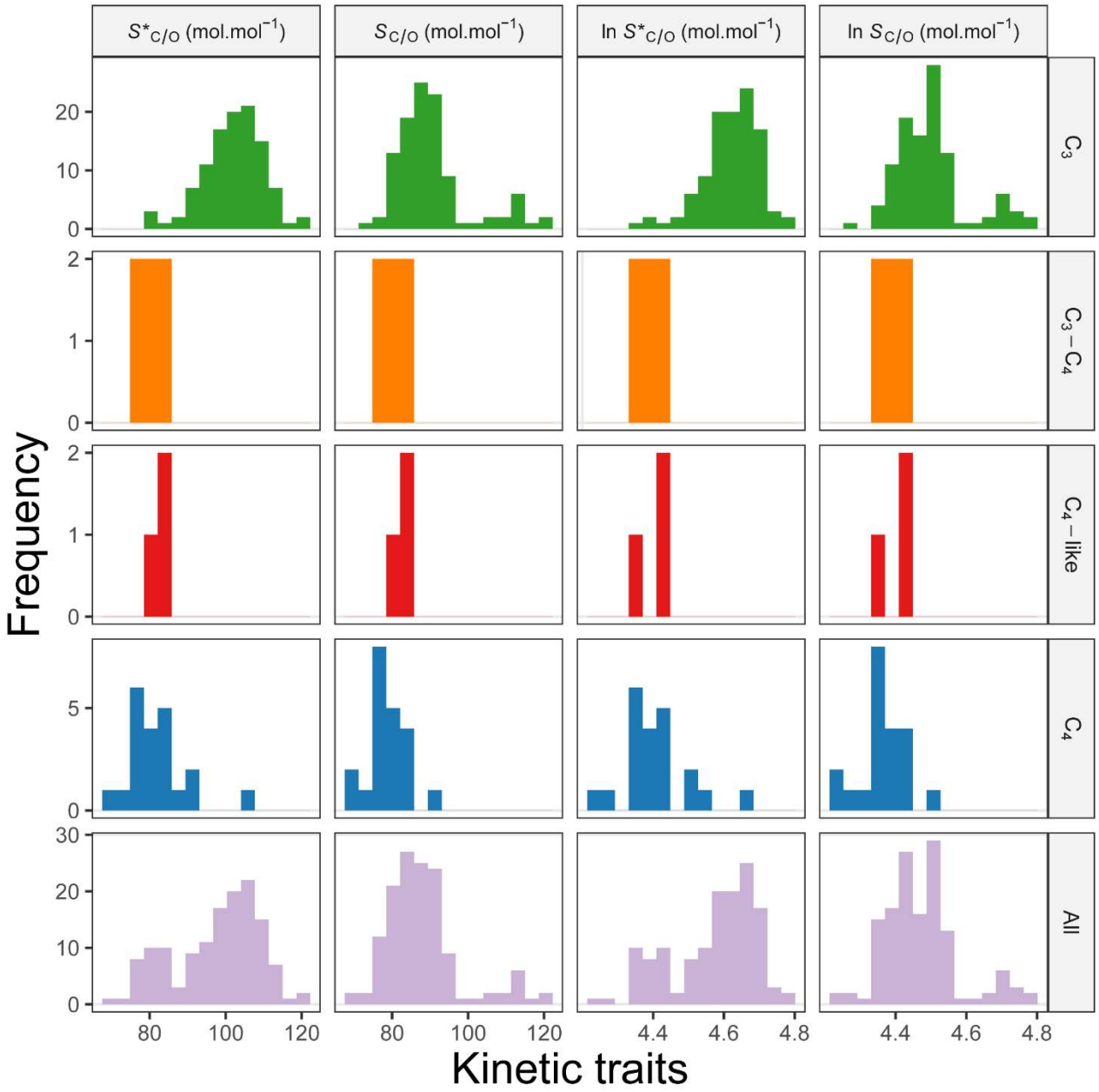
Appendix B, Figure S5. Heatmap depicting the normalised variation in rubisco kinetic traits across all studied photosynthetic organisms (± S.D. away from each respective kinetic trait mean).



Appendix B, Figure S6. The trends in relationships between all pairwise combinations of log transformed rubisco kinetic traits across all studied photosynthetic organisms. Data points are colour coded by taxonomic group (plant angiosperm rubisco: green, plant non-angiosperm rubisco: brown, and non-plant rubisco: blue).



Appendix B, Figure S7. Kinetic and phylogenetic constraints on rubisco adaptation across all studied photosynthetic organisms omitting known C_3-C_4, C_4-like and C_4 angiosperms **A)** Pairwise correlation coefficients (percent variance explained) and associated *p*-values between different rubisco kinetic traits assessed using non-phylogenetic least squares regression models or phylogenetic least squares regression models. Significance values are represented as α levels, where; α = 0.001 if $p < 0.001$, α = 0.01 if $0.001 < p < 0.01$, α = 0.05 if $0.01 < p < 0.05$, and α = ns if $p > 0.05$. **B)** The variation (%) in rubisco kinetic traits across photosynthetic organisms (omitting non-C_3 angiosperms) that can be explained by phylogenetic constraint and each catalytic trade-off. **C)** Boxplot of all variation explained in each kinetic trait by kinetic trait correlations in comparison to variation explained by phylogeny in all photosynthetic organisms. The phylogenetic constraints on the carboxylase-related traits Phy_{CX} (includes $Phy_{Sc/o}$, Phy_{Kcatc}, and Phy_{Kc}) and phylogenetic constraints on the oxygenase-related trait Phy_{ox} (includes Phy_{Ko} only) are presented separately.



Appendix B, Figure S8. The distributions of values for rubisco specificity in angiosperms before ($S^*_{C/O}$) and after ($S_{C/O}$) normalisation of measurements using the O_2 electrode method (Orr *et al.*, 2016; Prins *et al.*, 2016) relative to those quantified using high precision gas-phase-controlled ^{3}H-RuBP-fixation assays (Kane *et al.*, 1994). Species are grouped by their photosynthetic types (rows). Plants have been classified as those which perform C_3 photosynthesis (C_3; n = 107), C_4 photosynthesis (C_4; n = 6), C_3-C_4 intermediate (C_3-C_4; n = 3), C_4-like (C_4-like; n = 21). Both non-transformed and log-transformed non-normalised and normalised specificity values are shown, where respective units are shown in column labels.

B.3 Supplemental Tables

Appendix B, Table S1. The phylogenetic signal strength and associated significance level in rubisco kinetic traits across the phylogenetic tree in Appendix B, Figure S2 in five signal detection methods. Statistics are rounded to 3 decimal places, and significance values are represented as α levels, where; α = 0.001 if $p < 0.001$, α = 0.01 if $0.001 < p < 0.01$, α = 0.05 if $0.01 < p < 0.05$, and α = ns if $p > 0.05$.

Kinetic Trait	C mean		I		K		K *		Lambda	
	Stat	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***
$S_{C/O}$	0.502	0.001	0.516	0.001	0.001	0.05	0.001	0.05	0.873	0.001
k_{CatC}	0.343	0.001	0.303	0.001	0.002	0.001	0.003	0.001	0.940	0.05
K_C	0.344	0.001	0.339	0.001	0.002	0.01	0.002	0.001	0.948	0.001
K_C^{air}	0.295	0.001	0.291	0.001	0.002	0.001	0.002	0.001	0.52	0.05
K_O	0.125	0.05	0.142	0.05	0	ns	0	ns	0	ns

Appendix B, Table S2. Significance level of assessed differences in rubisco kinetic traits between C_3 and C_4 angiosperms compared when measured in the absence of the phylogenetic tree, and when correctly accounting for the phylogenetic non-independence of species. Significance values are represented as α levels, where; α = 0.001 if $p < 0.001$, α = 0.01 if $0.001 < p < 0.01$, α = 0.05 if $0.01 < p < 0.05$, and α = ns if $p > 0.05$

Kinetic Trait	***α***	
	– Phylogenetic information	**+ Phylogenetic information**
$S_{C/O}$	0.001	0.001
k_{CatC}	0.001	0.001
K_C	0.05	0.001
K_C^{air}	0.01	0.001
K_O	ns	ns

Appendix B, Table S3. The phylogenetic signal strength and associated significance level in rubisco kinetic traits across all studied photosynthetic organisms omitting known C_3-C_4, C_4-like and C_4 angiosperms using five different signal detection methods. Statistics are rounded to 3 decimal places and significance values are represented as α levels, where; α = 0.001 if $p < 0.001$, α = 0.01 if $0.001 < p < 0.01$, α = 0.05 if $0.01 < p < 0.05$, and α = ns if $p > 0.05$.

Kinetic	C mean		I		K		K *		Lambda	
Trait	*Stat*	*α*	*Stat*	*α*	*Stat*	*α*	*Stat*	*α*	*Stat*	*α*
$S_{C/O}$	0.717	0.001	0.245	0.001	0.006	0.001	0.003	0.001	1.009	0.001
k_{CatC}	0.53	0.001	0.313	0.001	0.002	0.001	0.001	0.01	0.955	0.001
K_C	0.75	0.001	0.309	0.001	0.005	0.001	0.002	0.001	0.986	0.001
K_C^{air}	0.657	0.001	0.327	0.001	0.005	0.001	0.002	0.001	0.983	0.001
K_O	0.332	0.001	0.151	0.01	0	0.05	0	ns	0.932	0.001

Appendix B, Table S4. Key of abbreviated species labels as used in Figure 2.2 and Figure 2.3A.

Code	Species
A.juv	*Aegilops juvenalis*
A.spe	*Aegilops speltoides*
A.vav	*Aegilops vavilovii*
A.squ	*Agriophyllum squarrosum*
A.sca	*Agrostis scabra*
A.sto	*Agrostis stolonifera*
A.cau	*Amaranthus caudatus*
A.hyb	*Amaranthus hybridus*
A.bra	*Amphicarpaea bracteata*
A.lat	*Arctagrostis latifolia*
A.myr	*Artemisia myriantha*
A.vul	*Artemisia vulgaris*
B.dis	*Brachypodium distachyon*
B.vul	*Beta vulgaris*
B.ano	*Bromus anomalus*
C.aru	*Calamagrostis arundinacea*
C.can	*Calamagrostis canescens*
C.fol	*Calamagrostis foliosa*
C.str	*Calamagrostis stricta subsp. inexpansa*
C.nut	*Calamagrostis nutkaensis*
C.cil	*Cenchrus ciliaris*
C.alb	*Chenopodium album*
C.mur	*Chenopodiastrum murale*
B.pet	*Blitum petiolare*
O.rub	*Oxybasis rubra*
C.ind	*Chrysanthellum indicum*
C.eci	*Citrullus ecirrhosus*
C.lan	*Citrullus lanatus*
D.dan	*Deschampsia danthonioides*
D.cin	*Desmodium cinereum*
D.int	*Desmodium intortum*
D.psi	*Desmodium psilocarpum*
E.far	*Elymus farctus*
E.tef	*Eragrostis tef*
E.fla	*Erythrina flabelliformis*
E.hel	*Euphorbia helioscopia*
E.mic	*Euphorbia microsphaera*
L.gig	*Lolium giganteum*
F.pra	*Festuca pratensis*
F.ano	*Flaveria anomala*
F.tri	*Flaveria trinervia*
F.bid	*Flaveria bidentis*
F.bro	*Flaveria brownii*
F.chl	*Flaveria chlorifolia*

F.cro	*Flaveria cronquistii*
F.flo	*Flaveria floridana*
F.koc	*Flaveria kochiana*
F.lin	*Flaveria linearis*
F.pal	*Flaveria palmeri*
F.pri	*Flaveria pringlei*
F.ram	*Flaveria ramosissima*
F.son	*Flaveria sonorensis*
F.vag	*Flaveria vaginata*
F.suf	*Flueggea suffruticosa*
F.vul	*Foeniculum vulgare*
G.can	*Glycine canescens*
G.max	*Glycine max*
H.vul	*Hordeum vulgare*
H.bra	*Hordeum brachyantherum*
H.mur	*Hordeum murinum*
L.pur	*Lablab purpureus*
L.cam	*Lepidium campestre*
L.artr	*Limonium artruchium*
L.gym	*Limonium gymnesicum*
L.mul	*Lolium multiflorum*
L.rig	*Lolium rigidum*
M.uni	*Macrotyloma uniflorum*
M.esc	*Manihot esculenta*
M.max	*Megathyrsus maximus*
M.ann	*Mercurialis annua*
M.vel	*Musa velutina*
N.tab	*Nicotiana tabacum*
xO.hyb	*Oryza barthii x Oryza glaberrima*
O.eic	*Oryza eichingeri*
O.mer	*Oryza meridionalis*
O.pun	*Oryza punctata*
O.sat	*Oryza sativa* subsp. *japonica*
P.ama	*Panicum amarum*
P.bis	*Panicum bisulcatum*
P.col	*Panicum coloratum*
P.deu	*Panicum deustum*
P.dic	*Panicum dichotomiflorum*
P.mid	*Panicum milioides*
P.mic	*panicum miliaceum*
P.mol	*Panicum monticola*
P.phr	*Panicum phragmitoides*
P.vir	*Panicum virgatum*
P.car	*Phaseolus carteri*
P.coc	*Phaseolus coccineus*
P.lun	*Phaseolus lunatus*
P.vul	*Phaseolus vulgaris*

P.pal	*Poa palustris*
P.dis	*Puccinellia distans*
P.lem	*Puccinellia lemmonii*
P.mar	*Puccinellia maritima*
P.nut	*Puccinellia nuttalliana*
P.mon	*Pueraria montana*
S.cer	*Secale cereale*
S.vir	*Setaria viridis*
S.ste	*Sphenostylis stenocarpa*
S.ole	*Spinacia oleracea*
S.lax	*Steinchisma laxum*
T.dic	*Triticum dicoccon*
T.mon	*Triticum monococcum*
T.tim	*Triticum timopheevii*
T.can	*Tephrosia candida*
T.pur	*Tephrosia purpurea*
T.rho	*Tephrosia rhodesica*
xTrit	*Secale cereale x Triticum aestivum*
T.aes	*Triticum aestivum*
T.aeg	*Triticum monococcum subsp. aegilopoides*
U.mos	*Urochloa mosambicensis*
U.pan	*Urochloa panicoides*
Z.may	*Zea mays*
O.sppA*	*Oryza* spp. Average Group A
A.sppB*	*Aegilops* Spp. Average Group B
L.sppC*	*Limonium* Spp. Average Group C
L.sppD*	*Limonium* Spp. Average Group D
L.sppE*	*Limonium* Spp. Average Group E

* Means have been taken between species which share identical *rbcL* sequences. *Oryza* spp. Average Group A include *O. longistaminata*, *O. glaberrima*, *O. sativa* f. *spontanea*, *O. sativa* subpsp. *indica* and *O. glumipatula*; *Aegilops* spp. Average Group B include *A. triuncialis*, *A. uniaristata*, *A. tauschii*, *A. comosa*, *A. biuncialis*, *A. cylindrica*; *Limonium* spp. Average Group C include *L. antonii-llorensii*, *L. gibertii*, *L. biflorum*; *Limonium* spp. Average Group D include *L. echioides*, *L. barceloi*, *L. balearicum*, *L. companyonis*; and *Limonium* spp. Average Group E include *L. ejulabilis*, *L. retusum*, *L. leonardi-llorensii*, *L. magallufianum*, *L. grosii*. **[*End of Figure Legend*].**

B.4 References

Gielly, L. & Taberlet, P. The use of chloroplast DNA to resolve plant phylogenies: Noncoding versus rbcL sequences. Molecular Biology and Evolution vol. 11 https://academic.oup.com/mbe/article-abstract/11/5/769/1008715 (1994).

APG. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. 181, 1–20 (2016).

Appendix C, Supplemental Data and Text

Supporting Chapter 3

Chapter 3, Contributors:

Bouvier, J.W., Emms, D.M. and Kelly, S.

Chapter 3, Author contributions:

J.W.B., D.M.E., and S.K. designed the research, J.W.B. and S.K. conceived the study, analysed the data, and wrote the manuscript, J.W.B. performed the research.

This thesis chapter has been published at the following location:

Bouvier, J.W., Emms, D.M. and Kelly, S. 2024. Proceedings of the National Academy of Sciences, 121(11), Article No e2321050121.

All supplemental data files are provided in the online version of this manuscript.

C.1 Supplemental Results

C.1.1 Form I rubisco ancillary chaperones exhibit a variable rate of molecular evolution

The molecular evolution of rubisco interacting partners were evaluated following the same method used to compute the percentile rank rate of evolution in *rbcL*/RbcL and *rbcS*/RbcS. These interacting partners included components involved in holoenzyme metabolic regulation (RUBISCO ACTIVASE (Rca)) as well as in holoenzyme folding and assembly (CHAPERONIN 10 (Cpn10), CHAPERONIN 20 (Cpn20), CHAPERONIN 60 (Cpn60), RBCX (RbcX), RUBISCO ACCUMULATION FACTOR 1 (Raf1), RUBISCO ASSEMBLY FACTOR 2 (Raf2)). This analysis revealed that the percentile rate of molecular evolution in rubisco chaperones was highly variable across all taxonomic groups, with no consistent pattern emerging in the relative extent of these genes compared to that experienced by

the cohort of all other genes encoded in each species genome (Appendix C, Figure S4A and S4B and table S11). Thus, similar to the case in the rubisco small subunit, rubisco chaperones do not experience the ubiquitous slow pace of molecular evolution observed in the large subunit of the holoenzyme.

C.1.2 Correlations in rubisco molecular and kinetic evolution are robust to biases in species sampling and model of sequence evolution

In order to evaluate the robustness of our integrated molecular and kinetic analysis, we conducted a systematic and comprehensive examination of possible methodological biases associated with our method to determine whether these may have impacted the results presented. Specifically, two such potential methodological issues were identified and include biases arising from incomplete species sampling in the rubisco kinetic dataset as well as biases arising from uncertainties or errors in the underlying phylogenetic gene tree that is used as the basis of the analysis. To address each of these in turn, a supporting set of additional analyses were designed. The results generated from these analyses were compared to those generated from the original analysis to assess the reliability of the presented conclusions.

First, we repeated the analysis of rubisco molecular and kinetic evolution but including only the minimum subset of species that captured most phylogenetic diversity across the species in the full analysis (see Methods, Appendix C, Figure S5). This additional analysis thus addresses any potential methodological artefacts caused by biases associated with incomplete species sampling (including overrepresentation of certain clades on the angiosperm phylogenetic tree of life).

Next, we repeated the analysis of rubisco molecular and kinetic evolution but using RbcL phylogenetic gene trees inferred from the same sequence alignment but using alternate models of sequence evolution. This analysis thus addresses any potential methodological artefacts associated with sources of error in the underlying phylogenetic tree caused by assuming a particular model of sequence evolution.

Analogous conclusions were generated from the results of each of these analyses (Appendix C, table S12 and S13) to those presented in the original investigation (Figure 3.4B). Thus, we confirm

that the results presented in our study are robust and are not an artefact caused by systematic biases in species sampling or caused by errors in phylogenetic reconstruction.

C.2 Supplemental Materials and Methods

C.2.1 Stratified sampling of rbcS sequences and phylogenetic tree inference

To account for potential biases in our analysis caused by some species exhibiting multiple copies of *rbcS*, random stratified sampling of the non-gapped *rbcS*/RbcS sequence alignments was conducted by species using 1,000 replicates with replacement. This process resulted in the generation of 1,000 unique *rbcS*/RbcS alignments for each taxonomic group, whereby each of these respective alignments contain only a single randomly selected copy of the *rbcS*/RbcS per species. In turn, each of these alignments were subject to data processing and phylogenetic tree inference using IQ-TREE (Nguyen *et al.*, 2015) following the method previously described.

C.2.2 Genomes and gene models

Complete sets of representative gene models for as many species in the rubisco sequence dataset as possible were acquired from either NCBI (https://www.ncbi.nlm.nih.gov/) or Phytozome V13 (Goodstein *et al.*, 2012). Where more than one such gene model resource was available for a given species, the most recent assembly version was chosen. In this way, complete sets of representative gene models were acquired for a total of 32 of the bacteria species, 27 of the land plant species, 8 of the SAR species, 6 of the red algae species and 4 of the green algae species analysed in the present study, respectively (Appendix C, table S1 and S7).

Predicted gene model sets were filtered to remove sequences with internal in-frame stop codons. Gene model sets were also filtered to keep only the longest gene model variant per gene. Moreover, owing to a lack of data availability of publicly available chloroplast or mitochondrial genomes for the eukaryotic species in the present analysis, and as organellar genomes contain fewer than 1% of genes encoded in the corresponding nuclear genome, only gene sequences encoded by the nuclear genomes of species in the land plant, green algae, red algae and SAR taxonomic groups were taken forward for analysis. Finally, after the above quality control checks were completed, a corresponding

proteome was generated from each species gene model set by *in silico* translation of the respective coding sequences.

C.2.3 Characterization of the set of enzymatic gene and protein sequences within orthogroups

The set of all genes within each species proteome that encode enzymes was determined using the DeepEC (Ryu, Kim and Lee, 2019) deep learning-based classifier algorithm. For this purpose, enzymes were defined as those protein sequences that could be assigned at least a partial enzyme commission (EC) number (i.e., at minimum, a single digit EC top-level code). On average 42.2% of all genes in the analysis encoded enzymes. A detailed breakdown of the metrics of enzyme ortholog pairwise comparisons for each species comparison can be found in ***Supplemental File 4 online***.

C.2.4 Identification and classification of Calvin-Benson-Bassham cycle enzyme isoforms in land plants

The set of genes which encode Calvin-Benson-Bassham cycle enzymes was first resolved in the model plant species *Arabidopsis thaliana*. To achieve this, the complete gene families to which each Calvin-Benson-Bassham cycle enzyme in *A. thaliana* belongs was determined based on available data in The Arabidopsis Information Resource (TAIR) database (http://arabidopsis.org) (Swarbreck *et al.*, 2008; Lamesch *et al.*, 2012). Following this, the photosynthetic isoforms in these gene families which are active in the Calvin-Benson-Bassham cycle in the chloroplast stroma were then identified based on several lines of evidence. 1) A high protein abundance based on whole-organism integrated protein abundance data obtained from the Protein Abundance Database (https://pax-db.org/) dataset 3702/323. 2) Leaf mRNA expression based on tissue-specific RNA sequencing data obtained from both the Arabidopsis eFP Browser V2.0 (http://bar.utoronto.ca/efp2/Arabidopsis/Arabidopsis_eFPBrowser2.html) and the EMBL-EBI (https://www.ebi.ac.uk/) dataset E-GEOD-53197. 3) Chloroplast-targeted protein subcellular localisation as predicted using both TargetP V2.0 (Emanuelsson *et al.*, 2000; Armenteros *et al.*, 2019) and Predotar V1.04 (Small *et al.*, 2004). 4) Gene orthology as inferred from trees generated for each Calvin-Benson-Bassham cycle gene family using IQ-TREE (Nguyen *et al.*, 2015). The

resulting set of photosynthetic isoforms encoding each Calvin-Benson-Bassham cycle enzyme in *A. thaliana* inferred from this multi-faceted analytical pipeline can be found in Appendix C, table S8.

The set of genes which encode the photosynthetic isoforms of Calvin Bensen Bassham cycle enzymes in all other 26 land plant species (apart from *A. thaliana*) for which genome sequence data was available in this study were then determined by orthology using data from the orthogroup inference analysis previously performed. Each group of orthologous protein sequences determined to encode a given Calvin-Benson-Bassham cycle enzyme across all species were aligned using the MAFFT L-INS-i algorithm (Katoh and Standley, 2013), and corresponding nucleotide coding sequence alignments were generated using PAL2NAL (Suyama, Torrents and Bork, 2006). Multiple sequence alignments were subject to the same data filtering and quality control criteria previously described to remove partial or incomplete sequences and subsequently delete any column positions which contain gaps. Finally, bootstrapped maximum likelihood phylogenetic trees were inferred by IQ-TREE (Nguyen *et al.*, 2015) following the method previously described. A similar analysis of Calvin-Bensen-Bassham cycle enzymes in other taxonomic groups was not able to be performed owing to a lack of the required data described here to determine the photosynthetic gene isoforms in these species.

C.2.5 Quantification of the relative extent of rubisco molecular evolution compared to all other Calvin-Benson-Bassham cycle enzymes in land plants

To determine the difference in evolutionary rate between rubisco and all other Calvin-Benson-Bassham cycle enzymes, the extent of molecular evolution measured between all other Calvin-Benson-Bassham cycle gene and protein orthologous sequences for each species pair (including *rbcS*/RbcS) were expressed as a percentage ratio of that measured in the corresponding *rbcL*/RbcL sequence. In cases where multiple percentage ratios are calculated for a given Calvin-Bensen-Bassham component in a given species pair (due to gene duplications, or due to a single species gene assembly matching multiple sub-species in the rubisco sequence dataset) the mean value was taken. The processed data generated from this analysis can be found in ***Supplemental File 5 online***.

C.2.6 Quantification of the percentile extent of rubisco chaperone molecular evolution within each taxonomic group

To evaluate the percentile rate of molecular evolution in the known chaperones of Form I rubisco in the context of all other genes in each taxonomic group, the exact same method was followed as previously described for *rbcL*/RbcL and *rbcS*/RbcS though the subject of the analysis was respectively altered. Here, for this investigation the putative set of genes which encode each Form I ancillary chaperone involved in holoenzyme metabolic regulation (RUBISCO ACTIVASE (Rca)) and in holoenzyme folding and assembly (BUNDLE SHEATH DEFECTIVE 2 (BSD2), CHAPERONIN 10 (Cpn10), CHAPERONIN 20 (Cpn20), CHAPERONIN 60 (Cpn60), RBCX (RbcX), RUBISCO ACCUMULATION FACTOR 1 (Raf1), RUBISCO ASSEMBLY FACTOR 2 (Raf2)) were first resolved in the model plant species *A. thaliana*. This was achieved using a previously published dataset (Gruber and Feiz, 2018) supplemented by information available in the TAIR database (http://arabidopsis.org) (Swarbreck *et al.*, 2008; Lamesch *et al.*, 2012). The resulting set of *Arabidopsis* chaperone genes thus identified can be found in Appendix C, table S9. Following this step, the corresponding set of genes encoding rubisco chaperones in all other species for which a complete gene assembly could be obtained were inferred using data from a separate OrthoFinder run performed with identical settings as previously described, but based on the translated proteomes of all organisms across all taxonomic groups together.

After the cohort of Form I rubisco chaperone genes were identified in all species, the percentile rates of nucleotide and protein evolution in these genes were calculated between each pairwise combination of species relative to all other pairs of orthologous sequences using the identical measurements previously generated from the analysis of *rbcL*/RbcL and *rbcS*/RbcS. In this way, analysis of some chaperones were omitted in certain taxonomic groups owing to the data quality and filtering steps that were performed as previously described. In cases where multiple percentiles are calculated for a chaperone in a given species pair (due to gene duplications, or due to a single species gene assembly matching multiple sub-species in the rubisco sequence dataset) the mean percentile was taken as previously described. The processed data quantifying the relative percentile

extent of molecular evolution in rubisco chaperones in the context of all other genes and proteins can be found in ***Supplemental File 5 online*** and ***Supplemental File 6 online***.

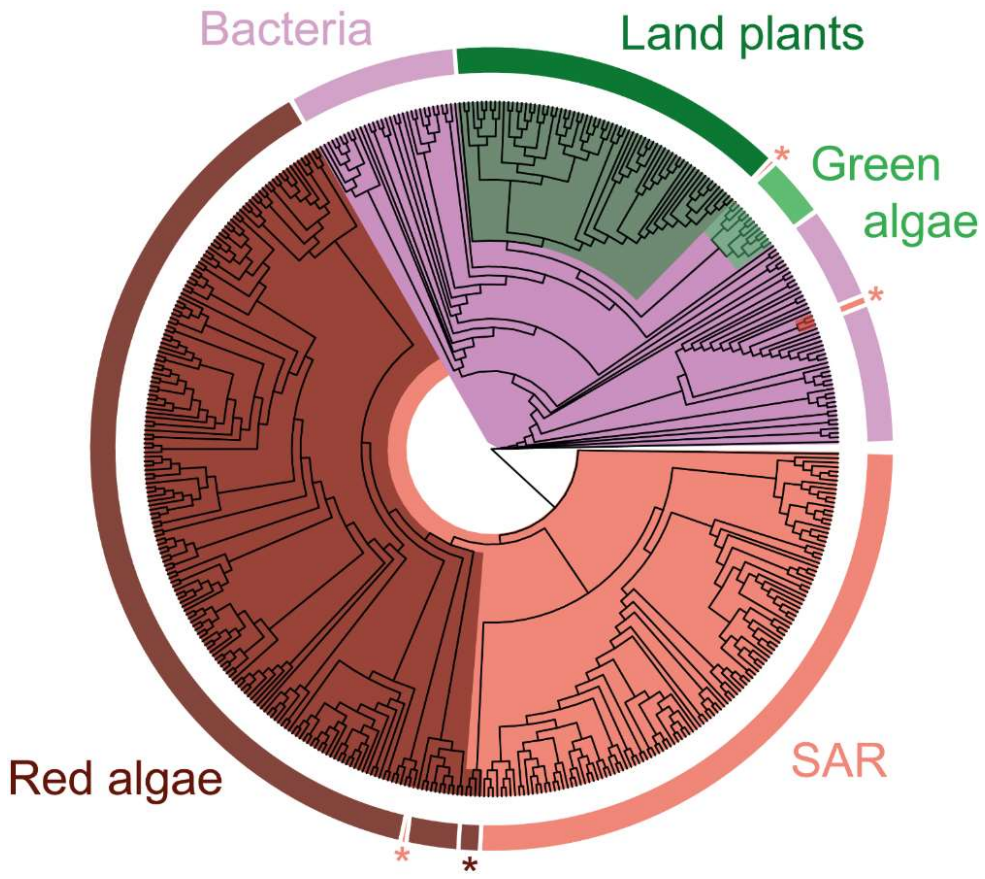
C.2.7 Accounting for potential errors in the integrated analysis of rubisco molecular and kinetic evolution associated with species sampling and model of sequence evolution

To control for potential errors associated with species sampling and overrepresentation of certain groups in the integrated molecular and kinetic investigation of rubisco evolution, an identical analysis was performed but including only the minimal subset of studied species which captured the majority of phylogenetic diversity (PD) in the kinetic dataset. For this purpose, the minimal subset of phylogenetically diverse studied species was identified by employing the Phylogenetic Diversity Analyzer V1.0.3 software (Chernomor *et al.*, 2015) with the 'greedy' algorithm. Here, PD is defined as the total tree length (i.e., the combined sum of all internal and terminal branch lengths) of the pruned phylogeny comprising a selected subset of sampled species. In brief, the unrooted RbcL phylogenetic tree of the 93 C_3 species was subject to systematic interrogation to identify the optimal combination of species at each iterative tree size (n = 2 – 93 species) which maximizes the PD score. Based on the results of this analysis, it was observed that PD plateaued at a threshold of 50 species (accounting for 54.8% of the total 93 C_3 species in the kinetic dataset) (Appendix C, Figure S5). The optimal composition of species at this respective threshold used for the basis of subsequent analyses included 31 dicotyledonous individuals and 19 monocotyledonous individuals, and are listed in Appendix C, table S10.

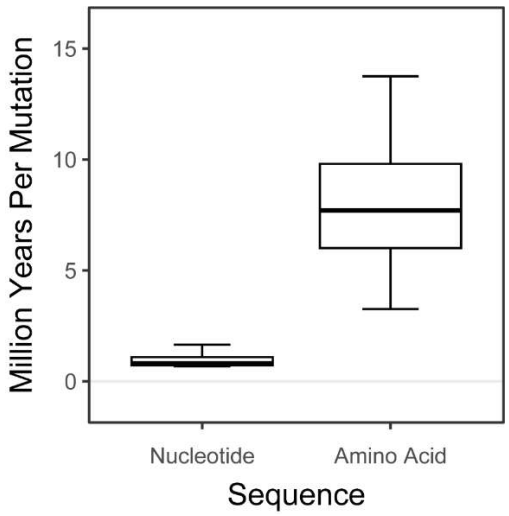
To control for potential artefacts associated with errors or uncertainties in phylogenetic tree inference in the integrated molecular and kinetic investigation of rubisco evolution, an identical analysis was performed using the complete set of species in the kinetic dataset but based on analogous gene trees. Specifically, these gene trees were generated using the same alignment of sequences and following the exact same method as previously described, but based on alternate best-fitting models of sequence evolution (LG+I+G4 and JTTDCMut+I+G4, respectively).

As the results of both of the above supplementary analyses based on the minimal subset of phylogenetically diverse species and based on alternate gene trees were identical to those generated from the original analysis, the conclusions in the present study were demonstrated to be valid and robust. Thus, the analyses presented are not an artefact caused by either systematic biases in species sampling or by errors in phylogenetic reconstruction.

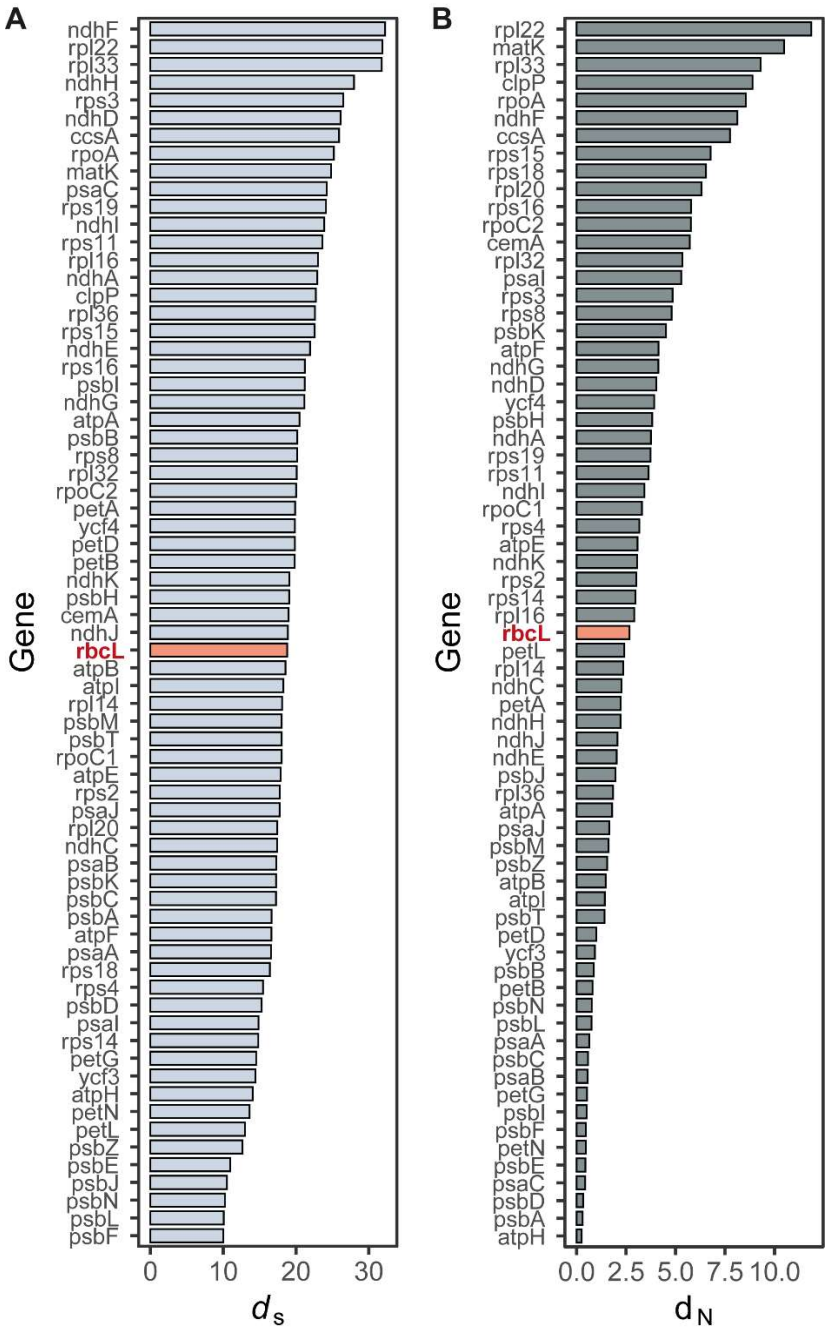
C.3 Supplemental Figures



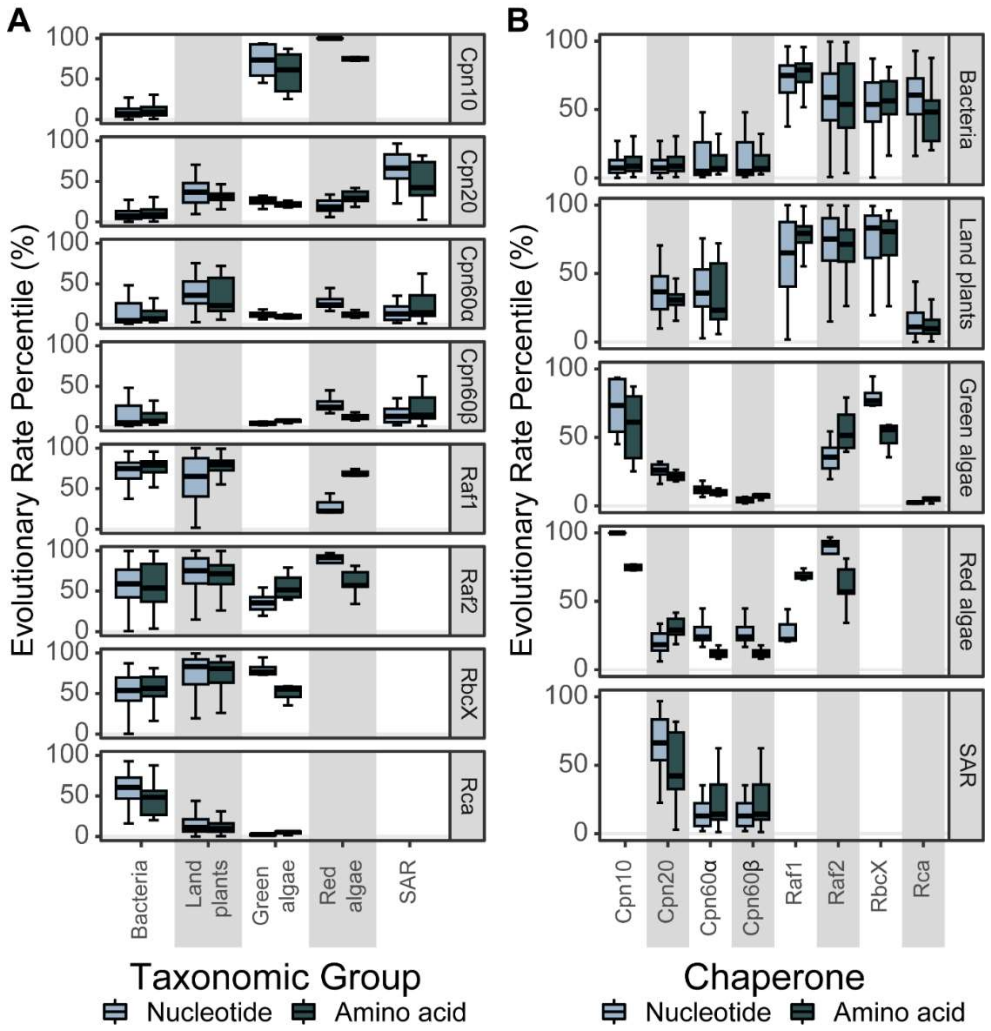
Appendix C, Figure S1. A phylogenetic tree of species in this study. Evolutionary history has been inferred from a multiple sequence alignment of the *rbcL* coding sequence in each species. The phylogeny is displayed as a cladogram for ease of visualisation, and species membership to different taxonomic groups (labelled) are highlighted by colour. Dark brown: red algae (*Rhodophyta*; *n* = 201). Light brown: SAR supergroup (*Stramenopiles*, *Alveolates*, and *Rhizaria*; *n* = 129). Lilac: bacteria (*Bacteria*; *n* = 78). Dark green: land plants (*Streptophyta*; *n* = 68). Light green: green algae (*Chlorophyta*; *n* = 12). Species at terminal nodes which cluster outside their designated taxonomic group are marked by an asterisk.



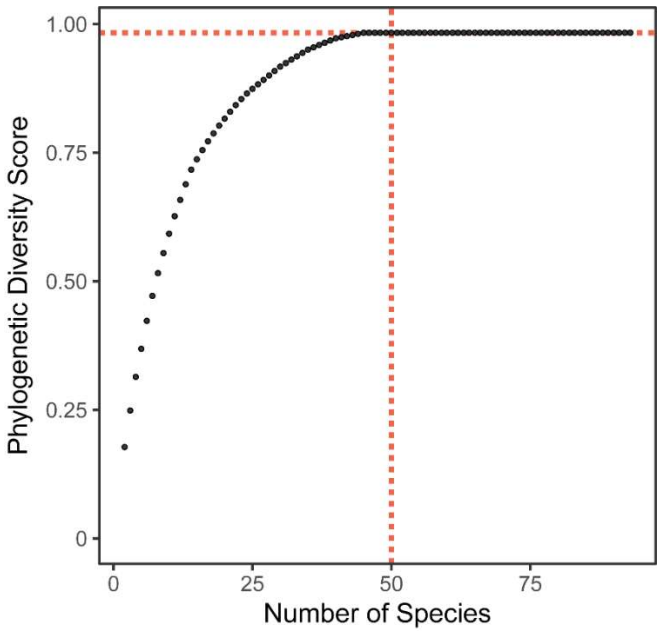
Appendix C, Figure S2. The rate of nucleotide and amino acid mutation in the rubisco large subunit experienced by C_3 angiosperms in the kinetic dataset, as expressed as the number of million years per individual sequence change. Molecular mutation rates were calculated since divergence from the last common ancestor at the base of the angiosperm clade 160 million years ago and considered an average *rbcL* of length 1428 nucleotide residues and 476 amino acid residues.



Appendix C, Figure S3. The extent of molecular evolution in rubisco in the context of the angiosperm plastid genome. A) Bar plot of the gene-wide rates of synonymous substitution (d_S) for all 69 ubiquitously conserved single-copy protein-coding genes in the angiosperm plastid genome. Genes are ordered in descending value of molecular evolution rate, and the relative position of *rbcL* is highlighted in red. B) As in (A), but for the gene-wide rates of nonsynonymous substitution (d_N) for all genes encoded in the angiosperm plastid genome. In this context, d_S and d_N are computed as the sum of all synonymous and non-synonymous mutations (substitutions per sequence site) which have occurred since the diversification of a conserved set of 773 angiosperms species. Figure redrawn from (Robbins and Kelly, 2023).



Appendix C, Figure S4. The extent of molecular evolution in each of rubisco's interacting chaperone partners. A) Boxplot of the extent of nucleotide and amino acid evolution (substitutions per sequence site) of each chaperone organised by taxonomic group. B) As in (A) but organised by gene/protein in alphabetical order. Chaperonin 60α: Cpn60α. Chaperonin 60β: Cpn60β. Chaperonin 10: Cpn10. Chaperonin 20: Cpn20. RbcX: RbcX. Rubisco accumulation factor 1: Raf1. Rubisco accumulation factor 2: Raf2. Rubisco activase: Rca. Although Bundle sheath defective 2 (Bsd2) is a known regulator of rubisco assembly, this chaperone was omitted from the analysis owing to insufficient sequence data. Data for this figure can be found in Appendix C, table S11.



Appendix C, Figure S5. The maximal phylogenetic diversity score which can be obtained at each iterative sample size of species in the rubisco kinetic dataset (*n* = 2 – 93 species). Phylogenetic diversity is defined as the total tree length (i.e., the combined sum of all internal and terminal branch lengths) of the phylogenetic tree inferred for each optimal subset of species at each sample size. The sample size (*n* = 50) at which phylogenetic diversity plateaus is indicated by a red dashed line. The optimal species set at this threshold are shown in Appendix C, table S10.

C.4 Supplemental Tables

Appendix C, Table S1. Summary of the sequence dataset used for analysis of rubisco molecular evolution in this study.

Taxonomic Group	*n*			
	Species	**Species Representative Gene Models**	***rbcL*/RbcL Sequences**	***rbcS*/RbcS Sequences**
Bacteria	78	32	78	86
Land plants	68	27	68	172
Green algae	12	4	12	24
Red algae	201	6	201	595
SAR	129	8	129	283
Total	**488**	**77**	**488**	**1140**

Appendix C, Table S2. Summary of the percentile rate of nucleotide and protein evolution in each rubisco subunit in context of all other genes in each taxonomic group. The median, first quartile (Q1), third quartile (Q3), the interquartile range (IQR), the mean and the standard error (S.E.) are provided.

Taxonomic Group	Rubisco Subunit	Sequence	Percentile Rate					
			Median	Q1	Q3	IQR	Mean	S.E.
Bacteria	**RbcL**	**Nucleotide**	2.597	0.847	5.693	4.846	4.729	0.317
Land plants			0.057	0.016	0.162	0.146	0.357	0.100
Green algae			0.042	0.019	0.088	0.070	0.053	0.016
Red algae			0.043	0.030	0.192	0.162	0.132	0.040
SAR			0.575	0.269	2.005	1.736	1.653	0.527
Bacteria		**Protein**	2.652	0.952	5.129	4.177	3.993	0.246
Land plants			2.180	1.573	3.314	1.741	3.092	0.201
Green algae			0.983	0.478	1.703	1.224	1.178	0.344
Red algae			0.933	0.874	1.141	0.267	1.168	0.186
SAR			2.614	2.048	3.840	1.792	3.169	0.467
Bacteria	**RbcS**	**Nucleotide**	25.205	14.899	39.836	24.937	28.373	0.837
Land plants			64.057	52.487	76.510	24.023	61.857	1.089
Green algae			3.961	1.905	5.643	3.738	5.198	2.078
Red algae			1.256	1.010	1.920	0.910	1.381	0.201
SAR			5.185	1.106	10.724	9.618	7.636	1.532
Bacteria		**Protein**	41.927	24.678	52.779	28.101	39.030	0.842
Land plants			52.024	36.131	67.519	31.387	52.229	1.002
Green algae			14.338	10.363	17.657	7.293	14.623	2.020
Red algae			8.630	7.019	16.601	9.581	11.352	1.697
SAR			12.445	4.047	17.256	13.210	12.410	1.518

Appendix C, Table S3. As in Appendix C, table S2 but in context of all other enzyme-encoding genes in each taxonomic group.

Taxonomic Group	Rubisco Subunit	Sequence	Percentile Rate					
			Median	Q1	Q3	IQR	Mean	S.E.
Bacteria	**RbcL**	**Nucleotide**	3.090	0.993	7.016	6.023	5.573	0.365
Land plants			0.070	0.031	0.174	0.143	0.384	0.108
Green algae			0.057	0.053	0.061	0.008	0.057	0.002
Red algae			0.093	0.077	0.161	0.085	0.152	0.037
SAR			0.498	0.241	1.316	1.075	1.275	0.452
Bacteria		**Protein**	2.978	0.961	5.706	4.745	4.334	0.274
Land plants			2.688	1.868	4.146	2.278	3.802	0.234
Green algae			0.229	0.192	0.286	0.094	0.334	0.115
Red algae			0.512	0.395	0.859	0.464	0.886	0.251
SAR			2.286	1.667	3.104	1.437	2.728	0.441
Bacteria	**RbcS**	**Nucleotide**	30.412	18.002	46.773	28.771	32.932	0.944
Land plants			71.256	60.765	83.175	22.410	68.478	1.055
Green algae			2.874	1.022	5.656	4.634	4.282	1.799
Red algae			1.044	0.525	2.056	1.531	1.382	0.299
SAR			6.122	0.819	11.510	10.692	8.097	1.639
Bacteria		**Protein**	47.314	29.298	56.569	27.271	43.035	0.887
Land plants			61.809	45.382	76.873	31.491	60.920	1.016
Green algae			17.982	12.037	23.533	11.496	17.801	2.583
Red algae			11.933	8.110	21.443	13.334	15.126	2.461
SAR			11.738	3.844	22.623	18.780	14.133	1.911

Appendix C, Table S4. Summary of the rate of nucleotide and protein evolution in each Calvin-Benson-Bassham cycle enzyme as a percentage ratio (%) of that measured in the rubisco large subunit (*rbcL*/RbcL). The median, first quartile (Q1), third quartile (Q3), the interquartile range (IQR), the mean and the standard error (S.E.) are provided. RBCS: rubisco small subunit. PGK: phosphoglycerate kinase. GAPDH-A: glyceraldehyde-3-phosphate dehydrogenase A subunit. GAPDH-B: glyceraldehyde-3-phosphate dehydrogenase B subunit. TPI: triose phosphate isomerase. FBA: fructose-bisphosphate aldolase. FBP: fructose-1,6-bisphosphatase. TKL: transketolase. SBP: sedoheptulose-bisphosphatase. RPI: ribose 5-phosphate isomerase. RPE: ribulose-p-3-epimerase. PRK: phosphoribulokinase.

Enzyme	Sequence	% of *rbcL*/RbcL Molecular Evolution					
		Median	Q1	Q3	IQR	Mean	S.E.
RBCS	**Nucleotide**	336.230	303.768	465.258	161.490	14215.472	6977.094
PGK		194.398	154.133	260.690	106.557	10222.926	8653.225
GAPDH-A		211.457	181.128	268.329	87.201	12074.030	9018.351
GAPDH-B		146.379	112.891	298.844	185.954	3293.398	1821.487
TPI		216.343	173.907	284.861	110.954	5679.185	3315.866
FBA		221.158	168.605	302.040	133.435	10306.953	8831.565
FBP		292.943	235.488	399.049	163.561	10201.657	5517.012
TKL		239.675	202.167	303.532	101.365	9195.482	5239.744
SBP		232.694	204.043	305.004	100.962	4667.535	3302.289
RPI		426.887	268.135	694.738	426.602	15006.493	11123.542
RPE		204.113	181.546	283.966	102.419	5462.525	4008.166
PRK		223.418	159.157	288.776	129.619	3818.743	2612.882
RBCS	**Protein**	677.203	513.747	880.073	366.326	13630.350	6097.334
PGK		188.897	128.489	271.151	142.662	1478.183	1094.706
GAPDH-A		168.140	103.943	231.241	127.298	7539.044	5584.847
GAPDH-B		102.634	69.460	180.491	111.031	7341.694	5148.096
TPI		227.879	190.426	290.050	99.624	3948.902	2281.597
FBA		145.765	107.797	180.354	72.556	5257.912	5090.570
FBP		267.831	189.307	397.150	207.842	25149.320	19153.272
TKL		284.397	225.285	339.574	114.289	5429.319	3058.388
SBP		141.192	96.510	217.906	121.396	2784.952	2048.169
RPI		186.521	114.161	284.244	170.083	12075.717	11840.183
RPE		233.601	167.671	306.208	138.536	2686.190	1714.928
PRK		167.644	132.689	248.041	115.352	1892.755	1245.509

Appendix C, Table S5. One-Sample Wilcoxon Signed Rank Test to assess significant differences in the rate of nucleotide and protein evolution between the rubisco large subunit (*rbcL*/RbcL) and each Calvin-Benson-Bassham cycle enzyme. A non-parametric test was used as data failed to conform to normality (Shapiro-Wilk test; $p < 0.05$). Statistics are rounded to three decimal places and corrected significance values are represented as α levels, where; α = 0.001 if $P < 0.001$, α = 0.01 if $0.001 < P < 0.01$, α = 0.05 if $0.01 < P < 0.05$, and α = ns if $P > 0.05$. Calvin-Bensen-Bassham cycle enzymes/subunits are abbreviated following the convention in Appendix C, table S4.

Enzyme	**Nucleotide**		**Protein**	
	Statistic	***α***	**Statistic**	***α***
RBCS	2476425	0.001	2476417	0.001
PGK	51040	0.001	48937	0.001
GAPDH-A	55278	0.001	49267	0.001
GAPDH-B	2550	0.001	1689	0.01
TPI	59685	0.001	59570	0.001
FBA	8984	0.001	7727	0.001
FBP	60378	0.001	60261	0.001
TKL	60378	0.001	60378	0.001
SBP	59684	0.001	50188	0.001
RPI	37128	0.001	33896	0.001
RPE	45450	0.001	45030	0.001
PRK	50721	0.001	49118	0.001

Appendix C, Table S6. Mean values of the ratio of rubisco large to small subunit percentile rank rate of evolution (*rbcL* to *rbcS* and RbcL to RbcS, respectively) and associated variation (±1 S.E.) in each taxonomic group.

% Ratio	Taxonomic Group				
	Bacteria	**Land Plants**	**Green Algae**	**Red Algae**	**SAR**
rbcL* : *rbcS	30.7 ± 4.3	0.6 ± 0.2	3.1 ± 1.6	10.7 ± 2.8	70.9 ± 46.8
RbcL : RbcS	12.2 ± 0.7	8.3 ± 1.2	8.3 ± 2.1	14.7 ± 3.0	40.2 ± 7.2

Appendix C, Table S7. List of all species in each taxonomic group for which either a nuclear (land plants, green algae, red algae, SAR) or bacterial (bacteria) genome could be acquired.

Taxonomic Group	Species
Bacteria	Acaryochloris marina
	Acidithiobacillus ferrooxidans
	Allochromatium vinosum
	Anabaenopsis circularis
	Arthrospira platensis
	Aurantimonas manganoxydans
	Crocosphaera subtropica
	Gloeobacter kilaueensis
	Gloeobacter violaceus
	Gloeomargarita lithophora
	Halomicronema hongdechloris
	Hydrogenophaga pseudoflava
	Methylacidimicrobium cyclopophantes
	Methylacidimicrobium tartarophylax
	Methylacidiphilum fumariolicum
	Methylacidiphilum infernorum
	Microcystis aeruginosa
	Microcystis viridis
	Nocardia nova
	Nocardia seriolae
	Novimethylophilus kurashikiensis
	Phaeobacter gallaeciensis
	Phormidesmis priestleyi
	Planktothrix agardhii
	Prochlorococcus marinus
	Prochlorothrix hollandica
	Raphidiopsis brookii
	Synechococcus elongatus
	Thermosynechococcus elongatus
	Thermosynechococcus vulcanus
	Thioflexothrix psekupsii
	Trichormus variabilis
Land Plants	Aegilops tauschii
	Amaranthus hypochondriacus
	Arabidopsis thaliana
	Brassica napus
	Brassica oleracea
	Brassica rapa
	Camellia sinensis
	Capsicum annuum
	Cucumis sativus
	Dendrobium catenatum
	Glycine soja
	Gossypium hirsutum
	Hevea brasiliensis
	Hordeum vulgare
	Lactuca sativa
	Mucuna pruriens
	Nicotiana attenuata
	Oryza sativa
	Panicum virgatum

	Phaseolus vulgaris
	Salvia splendens
	Sorghum bicolor
	Spinacia oleracea
	Triticum aestivum
	Triticum turgidum
	Triticum urartu
	Zea mays
Green Algae	Botryococcus braunii
	Chromochloris zofingiensis
	Dunaliella salina
	Volvox carteri
Red Algae	Chondrus crispus
	Cyanidiococcus yangmingshanensis
	Galdieria sulphuraria
	Gracilariopsis chorda
	Porphyra umbilicalis
	Porphyridium purpureum
SAR	Aureococcus anophagefferens
	Ectocarpus siliculosus
	Fistulifera solaris
	Microchloropsis salina
	Nannochloropsis gaditana
	Phaeodactylum tricornutum
	Thalassiosira oceanica
	Thalassiosira pseudonana

Appendix C, Table S8. The gene loci encoding the photosynthetic isoforms of Calvin-Benson-Bassham cycle enzymes in *Arabidopsis thaliana*.

Enzyme	Unique ID	Gene name	Arabidopsis TAIR ID
PGK	PGK1	PHOSPHOGLYCERATE KINASE 1	AT3G12780
	PGK2	PHOSPHOGLYCERATE KINASE 2	AT1G56190
GAPDH-A	GAPA-1	GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE A SUBUNIT 1	AT3G26650
	GAPA-2	GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE A SUBUNIT 2	AT1G12900
GAPDH-B	GAPB	GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE B SUBUNIT	AT1G42970
TPI	TPI	TRIOSEPHOSPHATE ISOMERASE	AT2G21170
FBA	FBA1	FRUCTOSE-BISPHOSPHATE ALDOLASE 1	AT2G21330
	FBA2	FRUCTOSE-BISPHOSPHATE ALDOLASE 2	AT4G38970
FBP	FBP1	FRUCTOSE 1,6-BISPHOSPHATE PHOSPHATASE	AT3G54050
TKL	TKL1	TRANSKETOLASE 1	AT3G60750
SBP	SBP	SEDOHEPTULOSE-BISPHOSPHATAS	AT3G55800
RPI	RPI	RIBOSE 5-PHOSPHATE ISOMERASE	AT3G04790
RPE	RPE	D-RIBULOSE-5-PHOSPHATE-3-EPIMERASE	AT5G61410
PRK	PRK	PHOSPHORIBULOKINASE	AT1G32060

Appendix C, Table S9. The gene loci encoding the chaperones involved in Form I rubisco assembly and metabolic regulation in *Arabidopsis thaliana*.

Chaperone	Unique ID	Gene name	Arabidopsis TAIR ID
Bsd2	Bsd2	BUNDLE SHEATH DEFECTIVE 2	AT3G47650
Cpn60α	Cpn60α_1	CHAPERONIN-60ALPHA1	AT2G28000
	Cpn60α_2	CHAPERONIN-60ALPHA2	AT5G18820
Cpn60β	Cpn60β_1	CHAPERONIN-60BETA1	AT1G55490
	Cpn60β_2	CHAPERONIN-60BETA2	AT3G13470
	Cpn60β_3	CHAPERONIN-60BETA3	AT5G56500
	Cpn60β_4	CHAPERONIN-60BETA4	AT1G26230
Cpn10	Cpn10_1	CHLOROPLAST CHAPERONIN 10	AT2G44650
	Cpn10_2	GROES	AT3G60210
Cpn20	Cpn20	CHAPERONIN 20	AT5G20720
RbcX	RbcX_1	RBCX1	AT4G04330
	RbcX_2	RBCX2	AT5G19855
Raf1	Raf1_1	RUBISCO ACCUMULATION FACTOR-LIKE PROTEIN	AT5G28500
	Raf1_2	RUBISCO ACCUMULATION FACTOR 1	AT3G04550
Raf2	Raf2	RUBISCO ASSEMBLY FACTOR 2	AT5G51110
Rca	Rca	RUBISCO ACTIVASE	AT2G39730

Appendix C, Table S10. The subset of 50 C_3 angiosperms which captured the vast majority of phylogenetic diversity encapsulated in the full rubisco kinetic dataset, as used for the basis of the analysis in Appendix C, table S12.

Division	Species
Dicot	Agriophyllum squarrosum
	Amphicarpaea bracteata
	Artemisia myriantha
	Artemisia vulgaris
	Beta vulgaris
	Chenopodiastrum murale
	Chenopodium album
	Citrullus ecirrhosus
	Desmodium cinereum
	Desmodium intortum
	Erythrina flabelliformis
	Euphorbia helioscopia
	Euphorbia microsphaera
	Flaveria cronquistii
	Flaveria pringlei
	Flueggea suffruticosa
	Foeniculum vulgare
	Glycine canescens
	Lepidium campestre
	Macrotyloma uniflorum
	Manihot esculenta
	Mercurialis annua
	Nicotiana tabacum
	Oxybasis rubra
	Phaseolus coccineus
	Pueraria montana
	Sphenostylis stenocarpa
	Spinacia oleracea
	Tephrosia candida
	Tephrosia purpurea
	Tephrosia rhodesica
Monoocot	Aegilops comosa
	Aegilops speltoides
	Agrostis stolonifera
	Arctagrostis latifolia
	Brachypodium distachyon
	Bromus anomalus
	Calamagrostis stricta subsp. inexpansa
	Deschampsia danthonioides
	Elymus farctus
	Lolium giganteum
	Lolium multiflorum
	Musa velutina
	Oryza barthii x Oryza glaberrima
	Oryza eichingeri
	Oryza glaberrima
	Panicum milioides
	Poa palustris
	Puccinellia distans

Appendix C, Table S11. Summary of the percentile rate of nucleotide and protein evolution in each of rubisco's interacting chaperone partners in context of all other genes in each taxonomic group. Data have been calculated between each pairwise combination of species within a taxonomic group for which either a nuclear (land plants, green algae, red algae, SAR) or bacterial (bacteria) genome could be acquired. Statistics are rounded to three decimal places, and include the median, first quartile (Q1), third quartile (Q3), the interquartile range (IQR), the mean and the standard error (S.E.). Rows and column containing N/A values denote missing data (see Methods).

Taxonomic Group	Rubisco Subunit	Sequence	Percentile Rate					
			Median	Q1	Q3	IQR	Mean	S.E.
Bacteria	**Rca**	**Nucleotide**	60.600	46.645	72.651	26.006	62.542	2.860
Land plants			11.058	6.231	21.374	15.142	17.656	1.000
Green algae			2.531	1.746	2.833	1.087	3.023	0.854
Red algae			N/A	N/A	N/A	N/A	N/A	N/A
SAR			N/A	N/A	N/A	N/A	N/A	N/A
Bacteria		**Protein**	48.212	26.908	56.477	29.569	48.414	3.275
Land plants			9.685	5.996	16.183	10.187	15.384	0.921
Green algae			5.226	3.735	6.019	2.284	5.330	1.213
Red algae			N/A	N/A	N/A	N/A	N/A	N/A
SAR			N/A	N/A	N/A	N/A	N/A	N/A
Bacteria	**Cpn10**	**Nucleotide**	7.383	3.834	13.185	9.352	10.065	0.410
Land plants			N/A	N/A	N/A	N/A	N/A	N/A
Green algae			73.334	54.076	92.558	38.482	72.007	9.232
Red algae			99.916	99.895	99.927	0.032	99.906	0.017
SAR			N/A	N/A	N/A	N/A	N/A	N/A
Bacteria		**Protein**	8.805	5.116	15.393	10.277	10.959	0.341
Land plants			N/A	N/A	N/A	N/A	N/A	N/A
Green algae			61.114	34.853	79.986	45.133	57.826	11.191
Red algae			74.744	72.862	76.651	3.789	74.769	1.162
SAR			N/A	N/A	N/A	N/A	N/A	N/A
Bacteria	**Cpn20**	**Nucleotide**	7.383	3.834	13.185	9.352	10.065	0.410
Land plants			36.601	23.947	47.920	23.973	37.041	0.791
Green algae			26.103	22.948	30.081	7.133	25.633	2.448
Red algae			18.322	14.026	26.449	12.423	20.410	2.235
SAR			66.293	53.703	83.382	29.678	66.736	3.769
Bacteria		**Protein**	8.805	5.116	15.393	10.277	10.959	0.341
Land plants			30.637	27.030	34.859	7.828	31.244	0.402
Green algae			22.078	19.036	23.944	4.908	21.814	1.401
Red algae			29.005	25.746	37.267	11.521	31.175	1.994
SAR			42.253	32.649	73.838	41.189	49.615	4.107
Bacteria	**Cpn60α**	**Nucleotide**	4.971	2.428	26.083	23.655	13.896	1.686
Land plants			35.767	25.918	53.009	27.091	39.147	1.066
Green algae			11.153	9.382	14.049	4.667	11.803	1.736
Red algae			24.023	21.721	31.270	9.549	27.034	2.148
SAR			12.939	5.571	22.243	16.672	16.451	2.728
Bacteria		**Protein**	6.870	5.314	16.609	11.294	11.737	0.987
Land plants			23.233	16.681	57.310	40.628	31.709	1.151
Green algae			9.104	8.050	11.442	3.392	9.664	0.917
Red algae			12.200	9.516	14.251	4.735	12.153	0.758
SAR			14.277	10.334	35.866	25.532	22.250	2.899
Bacteria	**Cpn60β**	**Nucleotide**	4.971	2.428	26.083	23.655	13.896	1.686
Land plants			N/A	N/A	N/A	N/A	N/A	N/A
Green algae			4.213	2.687	6.059	3.372	4.254	0.841

Red algae			24.023	21.721	31.270	9.549	27.034	2.148
SAR			12.939	5.571	22.243	16.672	16.451	2.728
Bacteria		**Protein**	6.870	5.314	16.609	11.294	11.737	0.987
Land plants			N/A	N/A	N/A	N/A	N/A	N/A
Green algae			7.646	6.150	8.271	2.121	7.614	1.105
Red algae			12.200	9.516	14.251	4.735	12.153	0.758
SAR			14.277	10.334	35.866	25.532	22.250	2.899
Bacteria	**Raf1**	**Nucleotide**	74.861	62.447	82.097	19.650	71.816	1.259
Land plants			65.047	40.394	87.520	47.126	64.306	1.594
Green algae			N/A	N/A	N/A	N/A	N/A	N/A
Red algae			22.207	21.367	33.185	11.818	28.966	7.614
SAR			N/A	N/A	N/A	N/A	N/A	N/A
Bacteria		**Protein**	78.876	70.092	83.569	13.477	76.585	1.079
Land plants			79.334	72.528	84.471	11.944	76.991	0.898
Green algae			N/A	N/A	N/A	N/A	N/A	N/A
Red algae			67.652	66.648	70.873	4.225	69.130	2.549
SAR			N/A	N/A	N/A	N/A	N/A	N/A
Bacteria	**Raf2**	**Nucleotide**	58.842	42.203	76.289	34.086	56.800	1.781
Land plants			75.092	59.362	90.363	31.002	73.069	1.087
Green algae			35.517	27.328	42.532	15.205	35.756	5.167
Red algae			91.233	84.730	93.894	9.163	83.658	8.248
SAR			N/A	N/A	N/A	N/A	N/A	N/A
Bacteria		**Protein**	53.812	36.805	83.573	46.768	56.263	1.887
Land plants			71.317	58.666	81.904	23.238	70.344	0.957
Green algae			51.523	42.443	66.557	24.114	55.425	6.637
Red algae			56.872	55.599	73.289	17.691	60.219	8.129
SAR			N/A	N/A	N/A	N/A	N/A	N/A
Bacteria	**RbcX**	**Nucleotide**	53.789	41.228	69.727	28.499	54.376	1.488
Land plants			83.203	61.384	92.168	30.784	74.490	1.323
Green algae			76.765	73.652	82.446	8.795	77.523	4.688
Red algae			N/A	N/A	N/A	N/A	N/A	N/A
SAR			N/A	N/A	N/A	N/A	N/A	N/A
Bacteria		**Protein**	56.356	46.610	70.736	24.125	56.760	1.243
Land plants			80.612	63.306	88.452	25.146	74.871	0.968
Green algae			55.560	45.803	58.422	12.619	55.154	6.568
Red algae			N/A	N/A	N/A	N/A	N/A	N/A
SAR			N/A	N/A	N/A	N/A	N/A	N/A

Appendix C, Table S12. The pairwise correlation coefficient (% variation explained), associated significance and direction of association from the results of the linear regression analysis between the extent of RbcL protein evolution and each rubisco kinetic trait but when considering the minimum combination of species which optimize phylogenetic diversity of species in the kinetic dataset, as predicated on Appendix C, Figure S5 and table S10. Significance values are represented as α levels, where $\alpha = 0.001$ if $P < 0.001$, $\alpha = 0.01$ if $0.001 < P < 0.01$, $\alpha = 0.05$ if $0.01 < P < 0.05$, and α = ns if $P > 0.05$.

Kinetic Trait	% Var Explained	*α*	Direction
$S_{C/O}$	13.5	0.01	+
k_{catC}	7.5	0.05	+
K_{catC}/k_C	6.4	0.05	+
K_C	ns	ns	ns
K_C^{air}	ns	ns	ns
K_O	ns	ns	ns
K_C/K_O	ns	ns	ns

Appendix C, Table S13. The pairwise correlation coefficient (% variation explained), associated significance and direction of association from the results of the linear regression analysis between the extent of RbcL protein evolution and each rubisco kinetic trait, but from the analysis of independent phylogenetic trees inferred from the same set of species sequence data but using alternative models of sequence evolution. Significance values are represented as α levels, where α = 0.001 if P < 0.001, α = 0.01 if 0.001 < P < 0.01, α = 0.05 if 0.01 < P < 0.05, and α = ns if P > 0.05.

Tree Model	Kinetic Trait	% Var Explained	*α*	Direction
LG+I+G4	$S_{C/O}$	7.2	0.01	+
	k_{catC}	11.0	0.001	+
	K_{catC}/k_C	9.6	0.01	+
	K_C	ns	ns	ns
	K_C^{air}	ns	ns	ns
	K_O	ns	ns	ns
	K_C/K_O	ns	ns	ns
JTTDCMut+I+G4	$S_{C/O}$	7.5	0.01	+
	k_{catC}	5.1	0.05	+
	K_{catC}/k_C	12.3	0.001	+
	K_C	ns	ns	ns
	K_C^{air}	ns	ns	ns
	K_O	ns	ns	ns
	K_C/K_O	ns	ns	ns

C.5 References

Armenteros, J.J.A. et al. (2019) 'Detecting sequence signals in targeting peptides using deep learning', Life Science Alliance, 2(5). Available at: https://doi.org/10.26508/lsa.201900429.

Chernomor, O. et al. (2015) 'Split diversity in constrained conservation prioritization using integer linear programming', Methods in Ecology and Evolution, 6(1), pp. 83–91. Available at: https://doi.org/10.1111/2041-210X.12299.

Emanuelsson, O. et al. (2000) 'Predicting subcellular localization of proteins based on their N-terminal amino acid sequence', Journal of Molecular Biology, 300(4), pp. 1005–1016. Available at: https://doi.org/10.1006/jmbi.2000.3903.

Goodstein, D.M. et al. (2012) 'Phytozome: A comparative platform for green plant genomics', Nucleic Acids Research, 40(D1). Available at: https://doi.org/10.1093/nar/gkr944.

Gruber, A.V. and Feiz, L. (2018) 'Rubisco assembly in the chloroplast', Frontiers in Molecular Biosciences, 5(MAR). Available at: https://doi.org/10.3389/fmolb.2018.00024.

Katoh, K. and Standley, D.M. (2013) 'MAFFT multiple sequence alignment software version 7: Improvements in performance and usability', Molecular Biology and Evolution. Oxford University Press, pp. 772–780. Available at: https://doi.org/10.1093/molbev/mst010.

Lamesch, P. et al. (2012) 'The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools', Nucleic Acids Research, 40(D1), pp. D1202–D1210. Available at: https://doi.org/10.1093/nar/gkr1090.

Nguyen, L.T. et al. (2015) 'IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies', Molecular Biology and Evolution, pp. 268–274. Available at: https://doi.org/10.1093/molbev/msu300.

Robbins, E.H.J. and Kelly, S. (2023) 'The Evolutionary Constraints on Angiosperm Chloroplast Adaptation', Genome Biology and Evolution, 15(6). Available at: https://doi.org/10.1093/gbe/evad101.

Ryu, J.Y., Kim, H.U. and Lee, S.Y. (2019) 'Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers', Proceedings of the National Academy of Sciences of the United States of America, 116(28), pp. 13996–14001. Available at: https://doi.org/10.1073/pnas.1821905116.

Small, I. et al. (2004) 'Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences', Proteomics, 4(6), pp. 1581–1590. Available at: https://doi.org/10.1002/pmic.200300776.

Suyama, M., Torrents, D. and Bork, P. (2006) 'PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments', Nucleic Acids Research, 34(WEB. SERV. ISS.), p. W609. Available at: https://doi.org/10.1093/nar/gkl315.

Swarbreck, D. et al. (2008) 'The Arabidopsis Information Resource (TAIR): Gene structure and function annotation', Nucleic Acids Research, 36(SUPPL. 1), pp. 1009–1014. Available at: https://doi.org/10.1093/nar/gkm965.

Appendix D, Supplemental Data and Text

Supporting Chapter 4

Chapter 4, Contributors:

Bouvier, J.W. and Kelly, S.

Chapter 4, Author contributions:

J.W.B. and S.K. conceived the study, designed the experiments, interpreted and analyzed the data, and wrote the manuscript. J.W.B. conducted the experiments.

This thesis chapter has been published as pre-print at the following location: Bouvier, J.W. and Kelly, S. 2024. bioRxiv, doi 10.1101/2024.08.29.610053

All supplemental data files are provided in the online version of this manuscript.

D.1 Supplemental Results

D.1.1 NOX expression results in tighter coordination between rates of assimilation and stomatal conductance

To further investigate the *iWUE* phenotype, the relative coordination between A and g_s was evaluated across the different light intensities. For this purpose, values of g_s were compared between those observed at $g_{s\ sat}$ and the modelled light intensity where 95% maximum assimilation ($g_{s\ A95}$) was achieved (beyond which neither CO_2 delivery nor light is considered limiting to photosynthesis). This revealed that rates of g_s in wild-type plants continue to increase by 8.1% after $g_{s\ A95}$ is achieved without resulting in any benefit in terms of CO_2 assimilation (Figure 4.3G, Appendix D, Figure S6A – S6C, table S3). In contrast, the magnitude of this "overshooting" behaviour was significantly dampened in transgenic plants, where an increase in g_s of only 3.3 – 4.4% was observed after $g_{s\ A95}$ was attained (Figure 4.3G, Appendix D, Figure S6A – S6C, table S3). This same result was

recapitulated when comparing the respective light intensities at which 95% maximum assimilation rate (PPFD $_{A95}$) and $g_{s\ sat}$ (PPFD $g_{s\ sat}$) were achieved. For instance, PPFD $_{A95}$ and PPFD $g_{s\ sat}$ were more similar to one another in transgenic (cyto-*Sm*NOX: 828.6 ± 161.3 µmol m^{-2} s^{-1}, 156.2%; stromal-*Sm*NOX: 621.4 ± 142.8 µmol m^{-2} s^{-1}, 124.6%; CIMS-*Sm*NOX: 847.6 ± 169.7 µmol m^{-2} s^{-1}, 168.2%) compared to wild-type plants (1385.7 ± 235.5 µmol m^{-2} s^{-1}, 290.0%) (Figure 4.3H, Appendix D, Figure S6D – S6F, table S3). Thus, the reduced stomatal density caused by *Sm*NOX expression resulted in an enhanced water-use efficiency and tighter coupling of g_s and *A* across different light conditions.

D.2 Supplemental Materials and Methods

D.2.1 Cloning and construct assembly

Cloning was performed using the Golden Gate one-step one-pot protocol (Engler *et al.*, 2009) with modified thermocycler conditions consisting of 80 cycles of 5 minutes digestion (37°C) and 5 minutes ligation (16°C). All reactions contained T4 DNA ligase (NEB, Catalog #M0202S) and the Type IIS restriction enzymes BsaI-HF®v2 (Level 1 assembly; NEB, Catalog #R3733S) or BpiI (Level 2 assembly; Thermo Fisher Scientific, Catalog # ER1012). Assembled plasmids were checked by colony PCR and gel electrophoresis using the OneTaq® 2X Master Mix with Standard Buffer (NEB, Catalog #M0482S) using generic primers that bind to the vector backbone, and sequences of inserts were confirmed by Sanger sequencing (Source BioScience, UK). Full plasmid sequencing (SNPsaurus, USA) was performed on all Level 2 constructs prior to transformation. Where 5' and 3' overhangs of vectors were not compatible for a given reaction, gene inserts were amplified by PCR with altered overhangs and resulting linear fragments were used directly for cloning. All PCR reactions for cloning was performed using Q5® High-Fidelity DNA Polymerase (NEB, Catalog #M0491S) following manufacturer's instructions for reaction set-up.

D.2.2 Vectors

pTEI071 and pJOG176 were a gift from Johannes Stuttmann (Addgene plasmid #105376; http://n2t.net/addgene:105376; RRID:Addgene_105376, and Addgene plasmid #105392; http://n2t.net/addgene:105392; RRID:Addgene_105392, respectively) (Gantner *et al.*, 2018). piCH41421 was a gift from Sylvestre Marillonnet and Nicola Patron (Addgene plasmid #50339;

http://n2t.net/addgene:50339; RRID:Addgene_50339) (Engler *et al.*, 2014). pICSL70008 was a gift from Nicola Patron (Addgene plasmid #50336; http://n2t.net/addgene:50336; RRID:Addgene_50336) (Engler *et al.*, 2014). pICH47742 and pICH49266 were a gift from Sylvestre Marillonnet (Addgene plasmid # 48001; http://n2t.net/addgene:48001; RRID:Addgene_48001, and Addgene plasmid #48024 ;http://n2t.net/addgene:48024; RRID:Addgene_48024) (Weber *et al.*, 2011). pAGM37443 was a gift from Sylvestre Marillonnet (Addgene plasmid #153218; http://n2t.net/addgene:153218; RRID:Addgene_153218) (Grützner *et al.*, 2021). EC15058 was a gift from Ben Miller.

D.2.3 Statistical analysis, model fitting, and parameter estimation of leaf gas exchange data

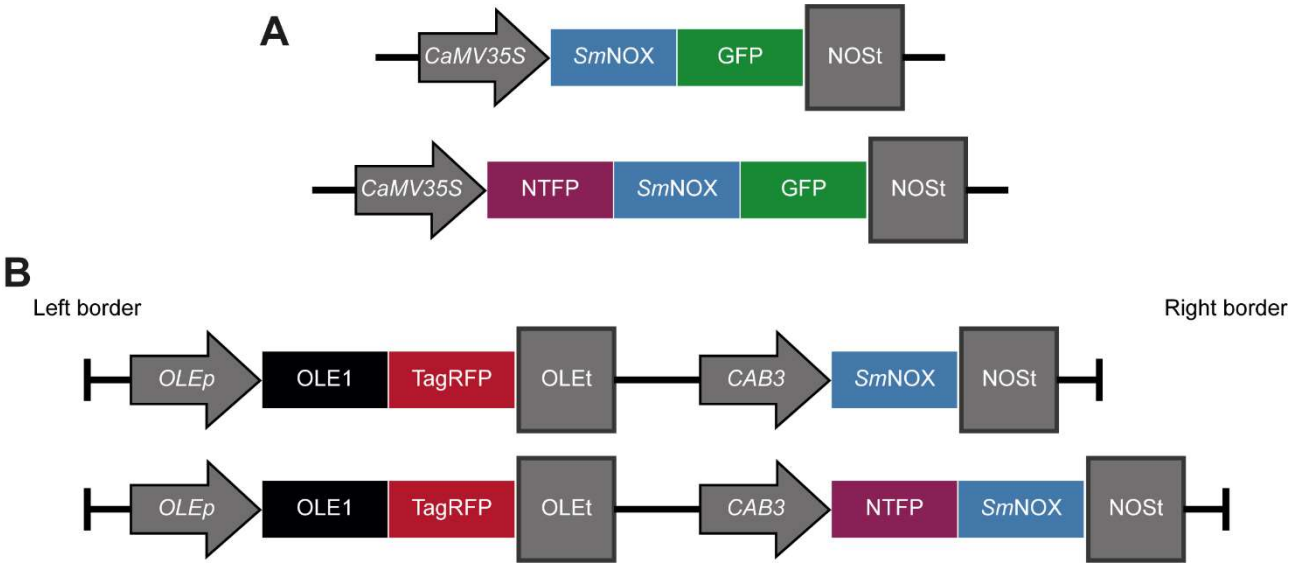
For statistical analysis of the response of plant gas exchange to light, non-linear curves were fit to the data by implementing the non-rectangular hyperbola model (Marshall and Biscoe, 1980) using the 'fit_photosynthesis' function in the R package photosynthesis V2.1.4 (Stinziano *et al.*, 2020, 2021). Values of light compensation point, mitochondrial respiration in the light, apparent quantum yield, and light-saturated electron transport rate were derived from these fitted models using this function. Light-saturated rates of assimilation (A_{sat}), stomatal conductance, and intrinsic water-use efficiency were manually determined as the highest value of these respective traits across the light response curve. The 95% maximum assimilation rate was also manually calculated from these models.

For statistical analysis of the response of plant gas exchange to CO_2, the Farquhar- von Caemmerer-Berry model of leaf photosynthesis was fit to the data using the 'fitaci' function from the R package plantecophys (Duursma, 2015). For this purpose, the bilinear method was used with triose phosphate utilization implemented to improve model fitting. Values of the maximum carboxylation rate of rubisco (V_{cmax}), maximum electron transport rate (J_{max}), CO_2 compensation point, the Michaelis constant of rubisco for CO_2 in 21% O_2 air, and the CO_2 concentration at the transition between rubisco-limited and ribulose-1,5-bisphosphate regeneration-limited photosynthesis were all derived from these fitted models using this function. Estimates of V_{cmax} and J_{max} re-scaled to 25°C were also generated from in-built temperature corrections of the data using the 'fitaci' function

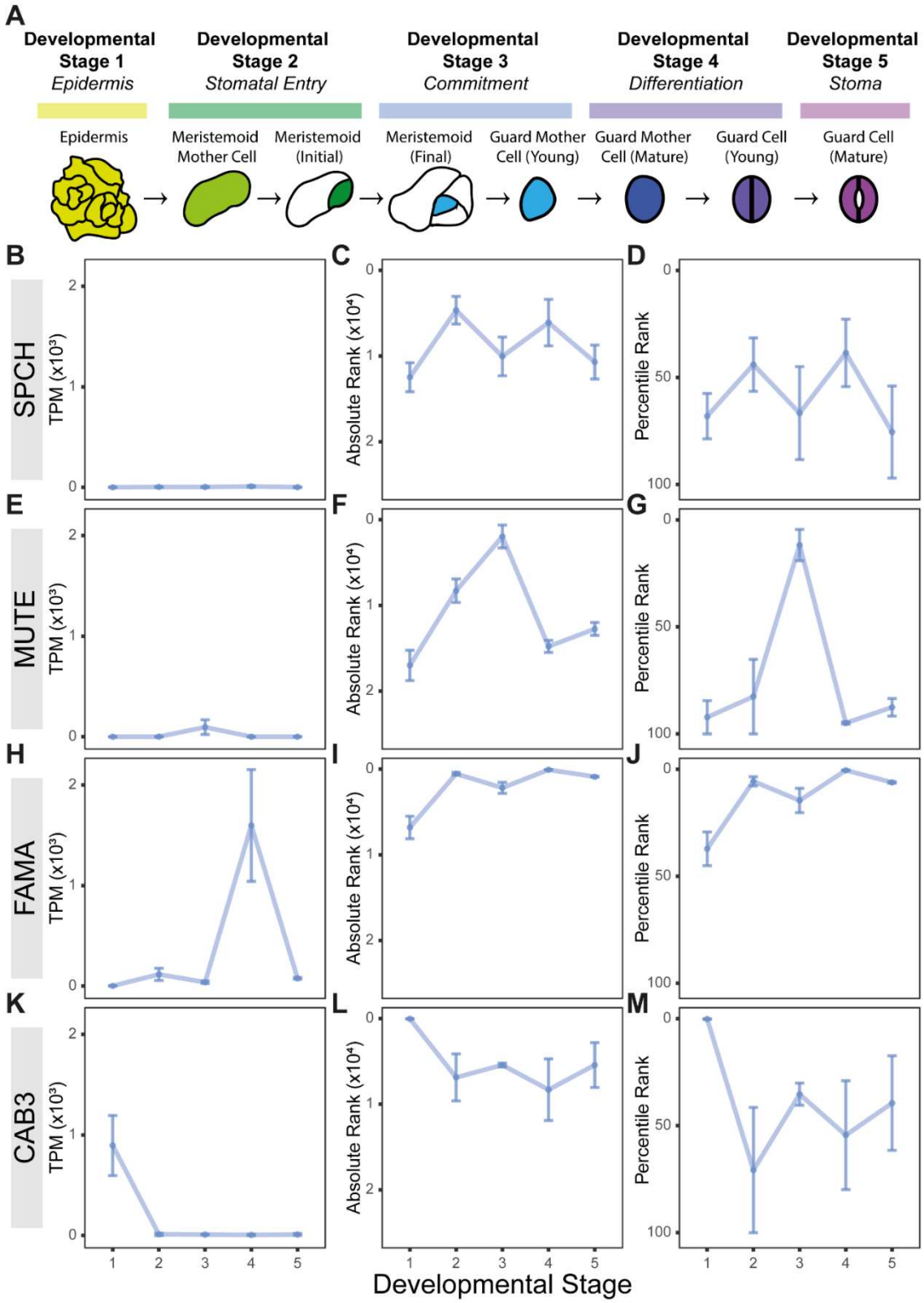
(Duursma, 2015). The CO_2–saturated assimilation rate (A_{max}) was manually determined as the highest value of this trait across the CO_2 response curve. The CO_2 concentration at the transition between RuBP regeneration-limited photosynthesis and triose-phosphate utilization-limited photosynthesis was not reported as could not be estimated in all cases.

Statistical analysis of stomatal kinetic parameters to step changes in light intensity between 100 μmol m^{-2} s^{-1} to 1000 μmol m^{-2} s^{-1} were derived using a dynamic sigmoidal mathematical model (Vialet-Chabrand, Dreyer and Brendel, 2013; McAusland *et al.*, 2016). For this purpose, model fitting was implemented using the licornetics V2.1.2 source code downloaded from the github repository https://github.com/lbmountain/licornetics. Estimates of the maximal rate of stomatal opening, time taken to achieve new steady-state stomatal conductance, and the initial lag in response time of g_s to the step increase in light intensity from 100 μmol m^{-2} s^{-1} to 1000 μmol m^{-2} s^{-1}, as well as the estimated time taken to achieve new steady-state stomatal conductance after the step decrease in light intensity from 1000 μmol m^{-2} s^{-1} to 100 μmol m^{-2} s^{-1} were all derived from models fitted in this way.

D.3 Supplemental Figures



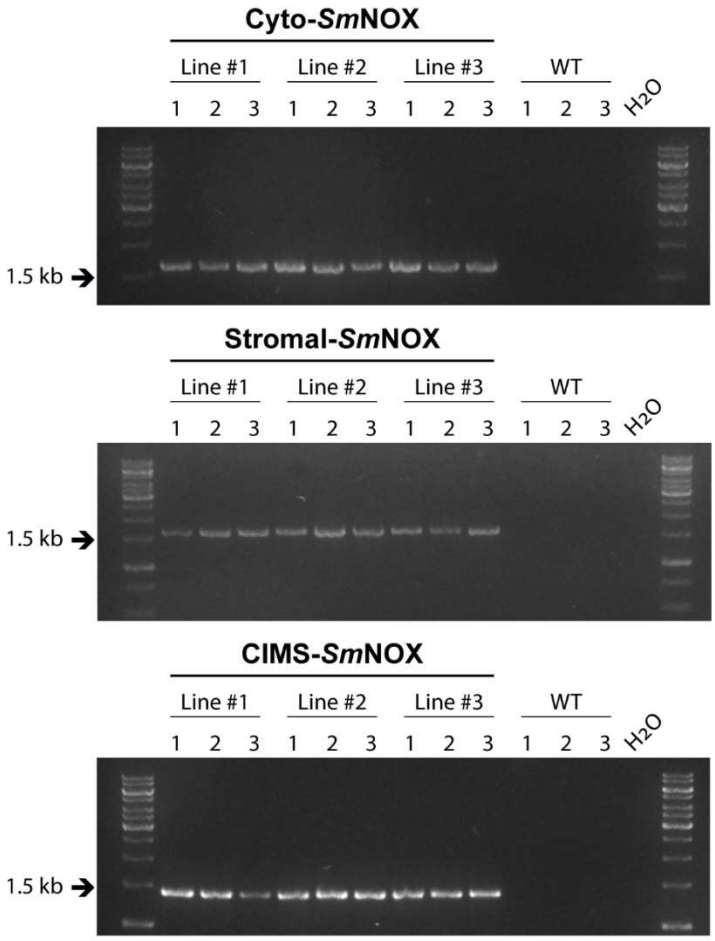
Appendix D, Figure S1. *Sm*NOX expression cassettes used for experiments in *Arabidopsis thaliana*. A) Simplified schematic of the expression cassette used for *Sm*NOX transient expression in *A. thaliana* protoplasts. The construct contains a CaMV 35S promoter (CaMV35S) and a nopaline synthase terminator (NOSt) controlling the expression of the *Sm*NOX coding sequence translationally fused to eGFP at the C-terminus. Top: untargeted *Sm*NOX. Bottom: *Sm*NOX targeted by an N-terminal fusion peptide (NTFP). NTFPs are derived from *A. thaliana* genes and include the predicted transit peptide from the photosystem II subunit QA gene or the full-length outer envelope protein 9 gene. B) As in (A), but a simplified schematic of the expression cassette used for *Sm*NOX stable expression in *A. thaliana*. In contrast to (A), *Sm*NOX does not contain a C-terminal eGFP and is driven under the control of the chlorophyll a/b-binding protein 3 (CAB3) promoter. All constructs also include an additional expression cassette encoding a seed coat specific fluorescent selectable marker which consists of monomeric TagRFP from *Entacmaea quadricolor* fused to the coding sequence of the *Arabidopsis thaliana* oleosin1 gene (OLE1) under the control of the native oleosin1 promoter (OLEp) and terminator (OLEt), respectively. The complete nucleotide and amino acid sequences for each expression cassette can be found in ***Supplemental File 2 online***.



Appendix D, Figure S2. The chlorophyll a/b binding protein (CAB3) promoter drives high gene expression in stomatal precursor cells. A) Schematic of the developmental trajectory of stomatal lineage differentiation including all major steps and fate transitions from undifferentiated epidermal cell (Developmental Stage 1), to stomatal lineage entry (Developmental Stage 2), stomatal commitment (Developmental Stage 3), stomatal differentiation (Developmental Stage 4), and finally, development of a fully-functional stoma (Developmental Stage 5). Figure adapted from (Adrian *et*

al., 2015). B) Temporal profile of SPEECHLESS (SPCH) transcript abundance (TPM; Transcripts Per Million) along the stomatal developmental trajectory in leaves of *Arabidopsis thaliana*. Respective developmental stages follow those described in (A). Representative transcriptomes of populations of cells at each developmental stage were obtained from (Adrian *et al.*, 2015). C) As in (B), but for the transcript abundance of SPCH expressed as a rank against the distribution of the transcript abundances of all other nuclear-encoded genes in *A. thaliana*. Ranks are calculated in descending order of gene transcript levels (i.e., low ranks denote genes with the highest relative expression and *vice versa*) and the y-axis has been inverted for ease of visualisation. D) As in (B), but for the transcript abundance of SPCH expressed as a percentile rank against the distribution of the transcript abundances of all other nuclear-encoded genes in *A. thaliana*. Gene percentile ranks are calculated in descending order of gene transcript levels (i.e., low percentile ranks denote genes with the highest relative expression and *vice versa*) and the y-axis has been inverted for ease of visualisation. E – G) As in (B), (C), and (D), but for MUTE. H – J) As in (B), (C), and (D), but for FAMA. K – M) As in (B), (C), and (D), but for CAB3. The raw data can be found in ***Supplemental File 5 online***. **[*End of Figure Legend*].**

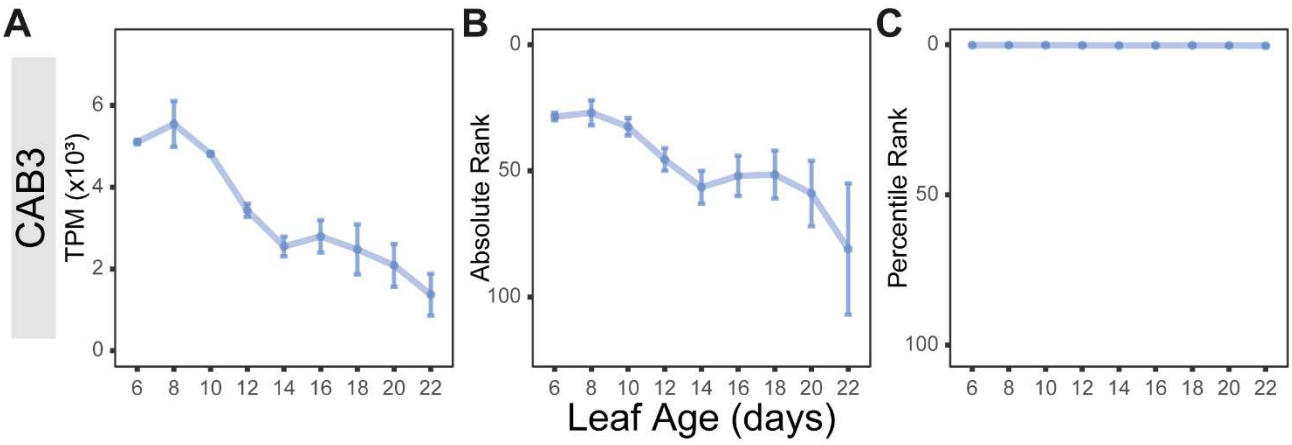
A



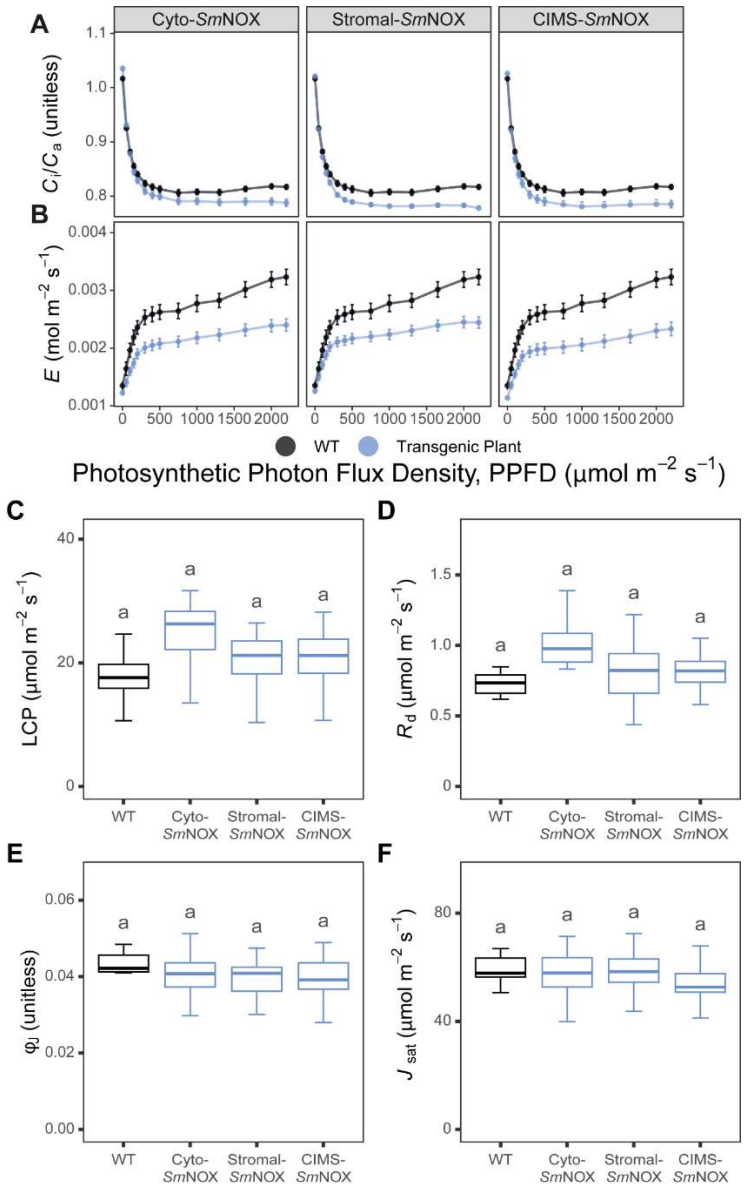
B

	Line	T_1 Seed			T_2 Seed
		TagRFP	WT	TagRFP:WT	TagRFP:WT
***Sm*NOX**	#1	62	23	**~3:1**	**100%**
	#2	120	43	**~3:1**	**100%**
	#3	254	84	**~3:1**	**100%**
Stromal *Sm*NOX	#1	54	20	**~3:1**	**100%**
	#2	70	30	**~3:1**	**100%**
	#3	55	18	**~3:1**	**100%**
RbcS–*Sm*NOX	#1	203	71	**~3:1**	**100%**
	#2	220	89	**~3:1**	**100%**
	#3	56	20	**~3:1**	**100%**
CIMS *Sm*NOX	#1	119	47	**~3:1**	**100%**
	#2	155	55	**~3:1**	**100%**
	#3	137	52	**~3:1**	**100%**

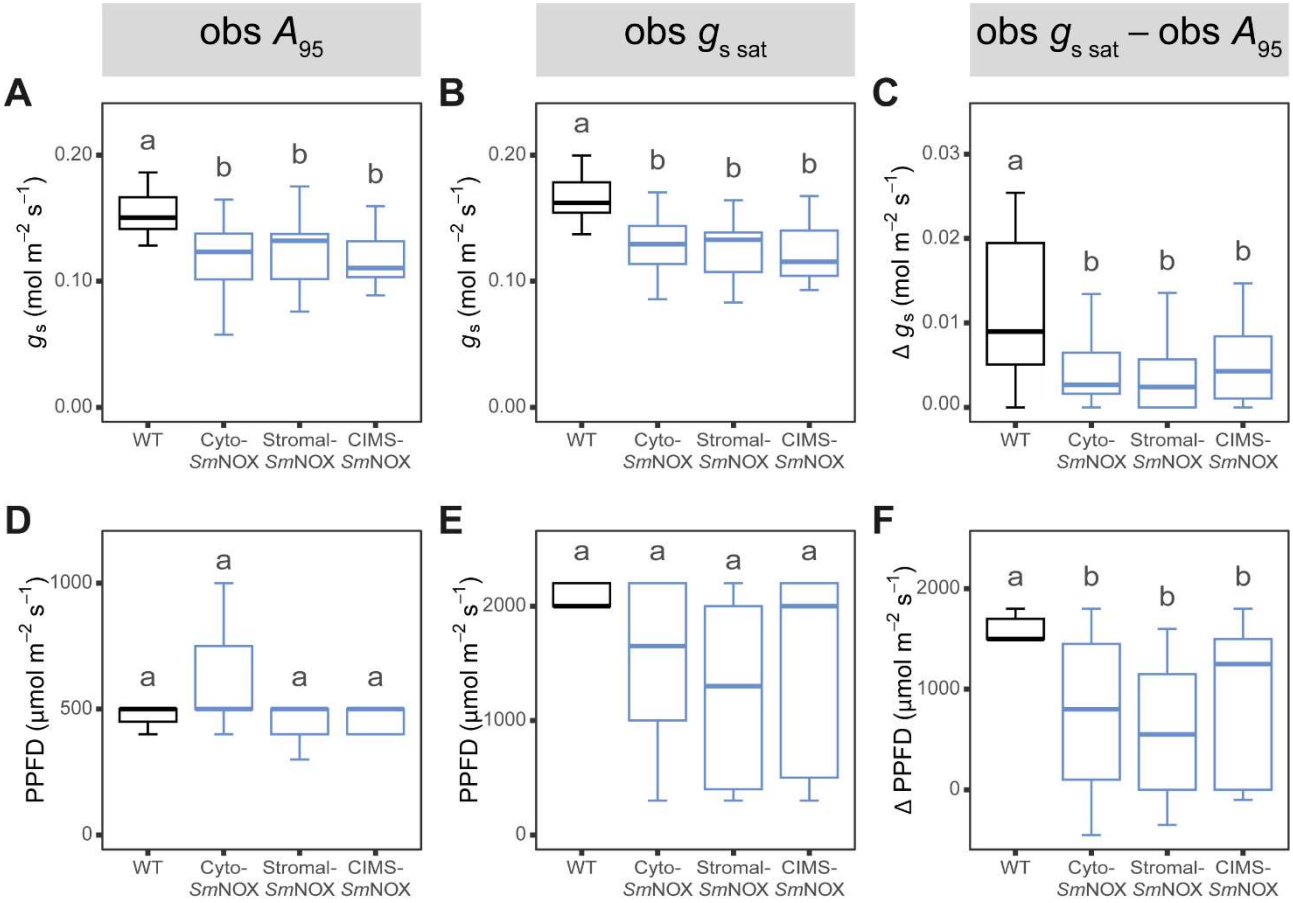
Appendix D, Figure S3. Verification of *Sm*NOX implementation. A) Genotyping of transgenic T_3 *Sm*NOX and wild-type (WT) plants based on primers specific for each gene insertion. The primers, reaction set up, and thermocycler conditions used for polymerase chain reaction (PCR) can be found in ***Supplemental File 3 online***. B) Segregation analysis data of seed from T_1 and T_2 generation plants.



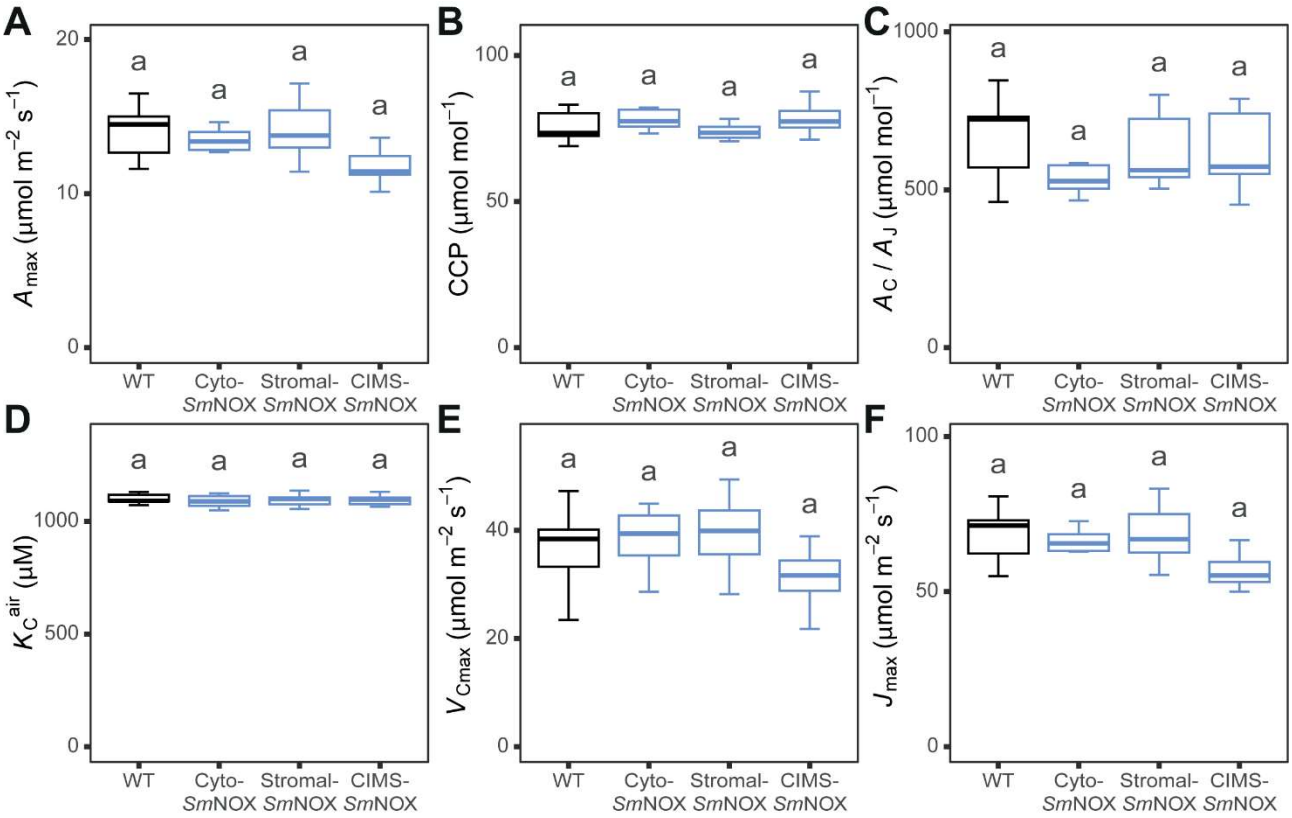
Appendix D, Figure S4. The chlorophyll a/b binding protein (CAB3) promoter drives highest levels of gene expression during early leaf development. A) Temporal profile of CAB3 transcript abundance (TPM; Transcripts Per Million) along a gradient of leaf aging in *Arabidopsis thaliana*. Representative transcriptomes from leaves at each age were obtained from (Woo *et al.*, 2016). C) As in (B), but for the transcript abundance of the CAB3 gene expressed as a rank against the distribution of transcript abundances of all other nuclear-encoded genes in *A. thaliana*. Gene ranks are calculated in descending order of gene transcript levels (i.e., low ranks denote genes with the highest relative expression and *vice versa*) and the y-axis has been inverted for ease of visualisation. D) As in (B), but for the transcript abundance of the CAB3 gene expressed as a percentile rank against the distribution of all other nuclear-encoded genes in *A. thaliana*. Gene percentile ranks are calculated in descending order of gene transcript levels (i.e., low percentile ranks denote genes with the highest relative expression and *vice versa*) and the y-axis has been inverted for ease of visualisation. The raw data can be found in ***Supplemental File 5 online***.



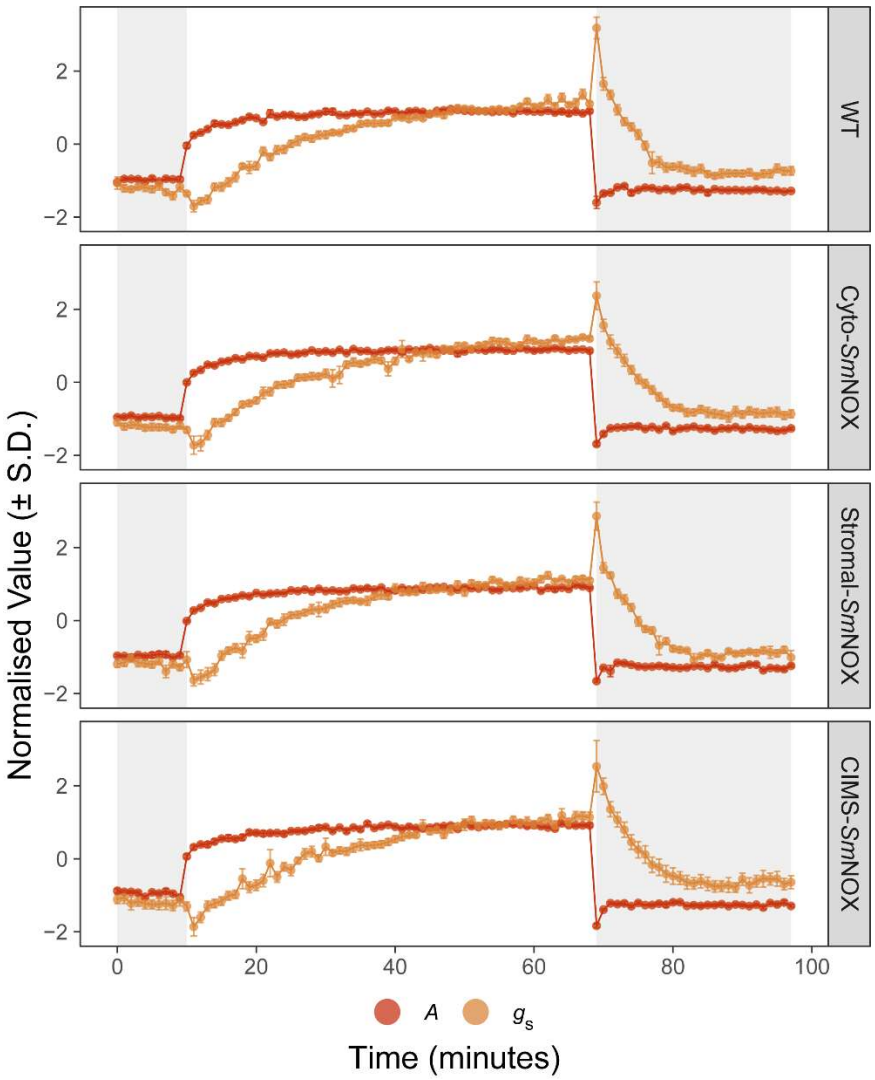
Appendix D, Figure S5. Additional parameters in the response of plant gas exchange to light (photosynthetic photon flux density: PPFD, µmol m^{-2} s^{-1}). A) The ratio of intercellular CO_2 to ambient CO_2 (C_i/C_a, unitless) at different PPFD in wild-type (black) and transgenic plants (blue). WT: wild-type. Cyto-*Sm*NOX: cytosolic *Sm*NOX. Stromal-*Sm*NOX: chloroplast stromal *Sm*NOX. CIMS-*Sm*NOX: chloroplast intermembrane space *Sm*NOX. Data represent mean ± 1 S.E. B) as in (A), but for plant transpiration rate (E, mol m^{-2} s^{-1}). C) Boxplot depicting the light compensation point (LCP, µmol mol^{-1}). The colour scheme and abbreviations follow that described in (A). D) As in (C), but for the mitochondrial respiration in the light (R_d, µmol m^{-2} s^{-1}). E) As in (C), but for the modelled apparent quantum yield (ΦJ, µmol m^{-2} s^{-1}). F) As in (C), but for the light-saturated electron transport rate (Jsat, µmol m^{-2} s^{-1}). All data represent an average across three independent single copy lines (n = 7 – 8 per individual line). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). The raw data can be found in ***Supplemental File 4 online***.



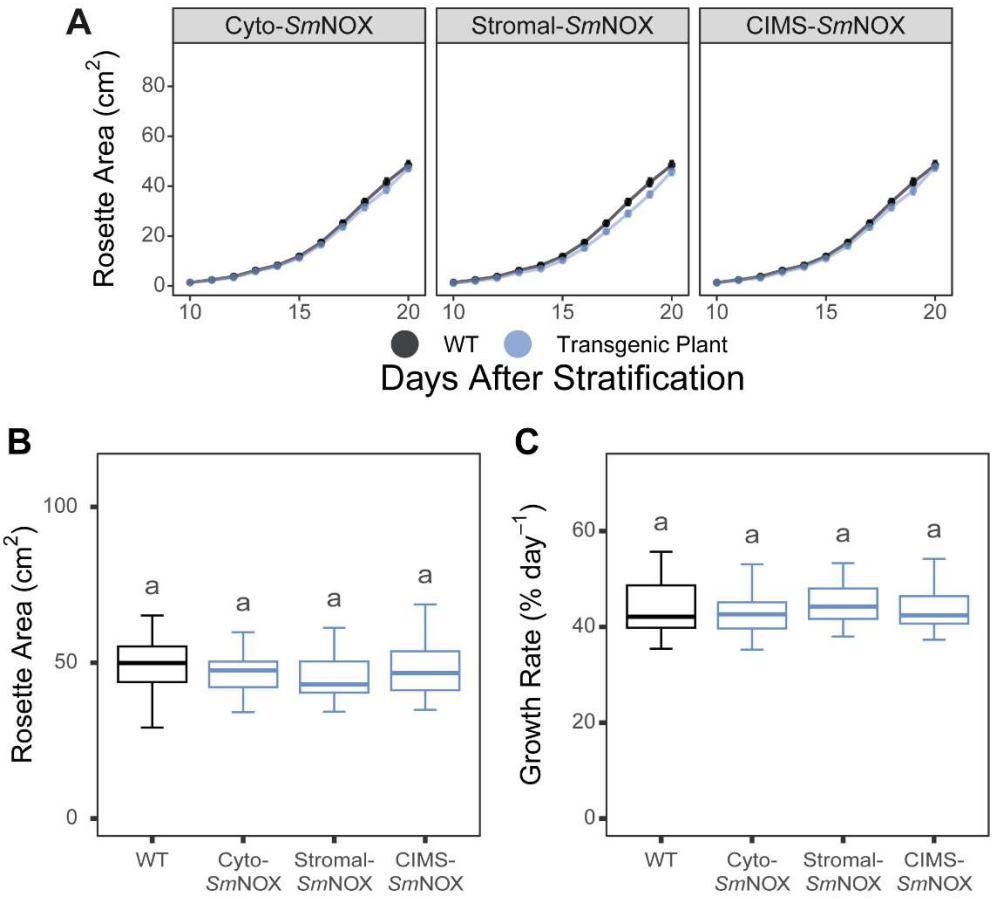
Appendix D, Figure S6. The relative coordination between rates of plant photosynthesis and gas exchange across light response curves. A) Boxplot depicting the stomatal conductance at the light intensity at which 95% maximum assimilation rate is achieved ($g_{s\ A95}$, mol m^{-2} s^{-1}) in wild-type (black) and transgenic plants (blue). WT: wild-type. Cyto-*Sm*NOX: cytosolic *Sm*NOX. Stromal-*Sm*NOX: chloroplast stromal *Sm*NOX. CIMS-*Sm*NOX: chloroplast intermembrane space *Sm*NOX. B) As in (A), but for the light-saturated stomatal conductance ($g_{s\ sat}$, mol m^{-2} s^{-1}). C) As in (A), but for difference in stomatal conductance between the light intensity at which 95% maximum assimilation rate and the light-saturated stomatal conductance is achieved (Δ g_s, mol m^{-2} s^{-1}). D) As in (A), but for the light intensity at which 95% maximum assimilation rate is achieved (PPFD $_{A95}$, µmol m^{-2} s^{-1}). E) As in (A), but for the light intensity at which light-saturated stomatal conductance is achieved (PPFD $_{gs\ sat}$, µmol m^{-2} s^{-1}). F) As in (A), but for the difference in irradiance between the light intensity at which 95% maximum assimilation rate and the light-saturated stomatal conductance is achieved (Δ PPFD, µmol m^{-2} s^{-1}). All data represent an average across three independent single copy lines (n = 7 – 8 per individual line). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). The raw data can be found in ***Supplemental File 4 online***.



Appendix D, Figure S7. The response of plant gas exchange to intercellular CO_2 concentration in wild-type (black) and transgenic plants (blue). WT: wild-type. Cyto-*Sm*NOX: cytosolic *Sm*NOX. Stromal-*Sm*NOX: chloroplast stromal *Sm*NOX. CIMS-*Sm*NOX: chloroplast intermembrane space *Sm*NOX. A) Boxplot depicting the CO_2–saturated assimilation rate (A_{max}, µmol m^{-2} s^{-1}). B) As in (A), but for the estimated CO_2 compensation point (CCP, µmol mol^{-1}). C) As in (A), but for the estimated CO_2 concentration at the transition between rubisco-limited and ribulose-1,5-bisphosphate (RuBP) regeneration-limited photosynthesis (A_C/A_J, µmol mol^{-1}). D) As in (A), but for the estimated Michaelis constant of rubisco for CO_2 in 21% O_2 air (K_C^{air}, µM). E) As in (A), but for the estimated maximum carboxylation rate of rubisco at 30°C (V_{cmax}, µmol m^{-2} s^{-1}). F) As in (A), but for the estimated maximum electron transport rate at 30°C (J_{max}, µmol e^- m^{-2} s^{-1}). All data represent an average across three independent single copy lines (n = 3 – 5 plants per line). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). The raw data can be found in ***Supplemental File 4 online***.



Appendix D, Figure S8. The normalised temporal response of plant CO_2 assimilation rate (*A*, µmol m^{-2} s^{-1}; red) and stomatal conductance (g_s, mol m^{-2} s^{-1}; orange) to step changes in light intensity. Shaded and unshaded areas represent periods of light intensity of 100 µmol m^{-2} s^{-1} and 1000 µmol m^{-2} s^{-1}, respectively. WT: wild-type. Cyto-*Sm*NOX: cytosolic *Sm*NOX. Stromal-*Sm*NOX: chloroplast stromal *Sm*NOX. CIMS-*Sm*NOX: chloroplast intermembrane space *Sm*NOX. For each parameter, data are normalised by the standard deviation away from the mean calculated across the entire temporal response of that individual. Data represent mean ± 1 S.E of normalised values. All data represent an average across three independent single copy lines (*n* = 3 – 4 per individual line). The raw data can be found in ***Supplemental File 4 online***.



Appendix D, Figure S9. The vegetative and reproductive performance of plants under normal watering conditions. (A) The increase in visible rosette area (cm^2) over time in wild-type (black) and transgenic plants (blue). WT: wild-type. Cyto-*Sm*NOX: cytosolic *Sm*NOX. Stromal-*Sm*NOX: chloroplast stromal *Sm*NOX. CIMS-*Sm*NOX: chloroplast intermembrane space *Sm*NOX. Data represent mean ± 1 S.E. B) Boxplot depicting the maximum visible rosette area at the end of vegetative growth (day = 20). The colour scheme and abbreviations follow that described in (A). C) As in (B), but for the average percentage growth rate of plants computed over the vegetative growth phase (% increase in rosette area day^{-1}). All data represent an average across three independent single copy lines (n = 10 per individual line for all measurements). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). The raw data can be found in ***Supplemental File 4 online***.

D.4 Supplemental Tables

Appendix D, Table S1. Predicted localisations of *Sm*NOX (including C-terminal GFP sequence) using target V2.0 software (Emanuelsson *et al.*, 2000; Armenteros *et al.*, 2019).

	Likelihood			
	Other	**Mitochondrial TP**	**Chloroplast TP**	**Thylakoid TP**
Cyto-*Sm*NOX	0.9948	0.0027	0	0

Appendix D, Table S2. Gas exchange parameters derived from light response (*A* – *PPFD*) curves. Values represent mean ± 1 S.E. All data represent an average across three independent single copy lines (*n* = 7 – 8 per individual line). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). Significant differences to WT are also highlighted in bold for ease of visualisation.

	WT	**Cyto-*Sm*NOX**	**Stromal-*Sm*NOX**	**CIMS-*Sm*NOX**
A_{sat} ($\mu mol\ m^{-2}\ s^{-1}$)	6.2 ± 0.3 [a]	5.3 ± 0.2 [a]	5.8 ± 0.2 [a]	5.2 ± 0.2 [a]
$g_{s\ sat}$ ($mol\ m^{-2}\ s^{-1}$)	0.166 ± 0.008 [a]	**0.126 ± 0.006 [b]**	**0.129 ± 0.005 [b]**	**0.121 ± 0.006 [b]**
$iWUE_{sat}$ ($\mu mol\ mol^{-1}\ H_2O$)	41.3 ± 1.4 [b]	**46.5 ± 1.6 [a]**	**48.5 ± 1.0 [a]**	**48.2 ± 2.2 [a]**
LCP ($\mu mol\ m^{-2}\ s^{-1}$)	17.7 ± 1.7 [a]	23.3 ± 1.7 [a]	18.5 ± 1.7 [a]	20.5 ± 1.3 [a]
R_d ($\mu mol\ m^{-2}\ s^{-1}$)	0.748 ± 0.084 [a]	0.941 ± 0.075 [a]	0.746 ± 0.074 [a]	0.783 ± 0.048 [a]
ΦJ (unitless)	0.043 ± 0.002 [a]	0.041 ± 0.002 [a]	0.040 ± 0.001 [a]	0.040 ± 0.002 [a]
J_{sat} ($\mu mol\ m^{-2}\ s^{-1}$)	59.2 ± 2.1 [a]	57.7 ± 1.9 [a]	57.8 ± 1.6 [a]	54.0 ± 1.5 [a]

N.B. A_{sat}: Light-saturated assimilation rate. $g_{s\ sat}$: Light-saturated stomatal conductance. $iWUE_{sat}$: Light-saturated instantaneous water-use efficiency. LCP: Light compensation point. R_d: Mitochondrial respiration in the light. ΦJ: Modelled apparent quantum yield computed as the initial slope of the response of J to PPFD. J_{sat}: Light-saturated electron transport rate.

Appendix D, Table S3. Gas exchange parameters describing the relative coordination between plant photosynthesis and gas exchange derived from light response (*A* – *PPFD*) curves. Values represent mean ± 1 S.E. All data represent an average across three independent single copy lines (*n* = 7 – 8 per individual line). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). Significant differences to WT are also highlighted in bold for ease of visualisation.

	WT	Cyto-*Sm*NOX	Stromal-*Sm*NOX	CIMS-*Sm*NOX
$g_{s\ A95}$ (mol m^{-2} s^{-1})	0.154 ± 0.008 [a]	**0.120 ± 0.006 [b]**	**0.125 ± 0.005 [b]**	**0.116 ± 0.006 [b]**
$g_{s\ sat}$ (mol m^{-2} s^{-1})	0.166 ± 0.008 [a]	**0.126 ± 0.006 [b]**	**0.129 ± 0.005 [b]**	**0.121 ± 0.006 [b]**
Δ g_s (mol m^{-2} s^{-1})	0.012 ± 0.004 [a]	**0.005 ± 0.001 [b]**	**0.004 ± 0.001 [b]**	**0.005 ± 0.001 [b]**
PPFD $_{A95}$ (µmol m^{-2} s^{-1})	471.4 ± 18.4 [a]	695.2 ± 89.9 [a]	581.0 ± 71.6 [a]	642.9 ± 98.0 [a]
PPFD $_{gs\ sat}$ (µmol m^{-2} s^{-1})	1857.1 ± 245.8 [a]	1523.8 ± 157.9 [a]	1202.4 ± 160.2 [a]	1490.5 ± 175.9 [a]
Δ PPFD (µmol m^{-2} s^{-1})	1385.7 ± 235.5 [a]	**828.6 ± 161.3 [b]**	**621.4 ± 142.8 [b]**	**847.6 ± 169.7 [b]**

N.B. $g_{s\ A95}$: stomatal conductance at the light intensity at which 95% maximum assimilation rate is achieved. $g_{s\ sat}$: Light-saturated stomatal conductance. Δ g_s: The difference in stomatal conductance between the light intensity at which 95% maximum assimilation rate is achieved and the light-saturated stomatal conductance is achieved. PPFD $_{A95}$: Light intensity at which 95% maximum assimilation rate is achieved. PPFD $_{gs\ sat}$: Light intensity at which light-saturated stomatal conductance is achieved. Δ PPFD: The difference in irradiance between the light intensity at which 95% maximum assimilation rate is achieved and the light-saturated stomatal conductance is achieved.

Appendix D, Table S4. Gas exchange parameters derived from CO_2 response (A – C_i) curves. Values represent mean ± 1 S.E. All data represent an average across three independent single copy lines (n = 3 – 5 per individual line). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). Significant differences to WT are also highlighted in bold for ease of visualisation.

	WT	**Cyto-*Sm*NOX**	**Stromal-*Sm*NOX**	**CIMS-*Sm*NOX**
$\boldsymbol{A}_{\mathbf{max}}$ (μmol m^{-2} s^{-1})	13.6 ± 0.9 [a]	12.8 ± 0.9 [a]	14.2 ± 0.5 [a]	11.4 ± 0.6 [a]
CCP (μmol mol^{-1})	75.7 ± 1.6 [a]	79.7 ± 2.2 [a]	72.9 ± 1.4 [a]	78.3 ± 1.6 [a]
$\boldsymbol{K_C^{air}}$ (μM)	1100.4 ± 6.9 [a]	1089.2 ± 7.9 [a]	1092.7 ± 6.9 [a]	1093.8 ± 6.7 [a]
$\boldsymbol{V}_{\mathbf{c\ max}}$ (μmol m^{-2} s^{-1})	37.0 ± 2.6 [a]	37.3 ± 2.4 [a]	39.4 ± 1.9 [a]	31.4 ± 1.5 [a]
$\boldsymbol{V}_{\mathbf{c\ max}}^{\mathbf{25°C}}$ (μmol m^{-2} s^{-1})	21.2 ± 1.6 [a]	21.6 ± 1.4 [a]	22.8 ± 1.1 [a]	18.1 ± 0.9 [a]
$\boldsymbol{J}_{\mathbf{max}}$ (μmol m^{-2} s^{-1})	66.1 ± 4.4 [a]	62.8 ± 4.5 [a]	68.7 ± 2.7 [a]	55.0 ± 3.1 [a]
$\boldsymbol{J}_{\mathbf{max}}^{\mathbf{25°C}}$ (μmol m^{-2} s^{-1})	54.7 ± 3.7 [a]	52.0 ± 3.8 [a]	56.9 ± 2.2 [a]	45.6 ± 2.5 [a]
A_C/A_J (μmol mol^{-1})	657.6 ± 41.6 [a]	563.3 ± 27.8 [a]	622.7 ± 34.6 [a]	626.4 ± 38.2 [a]

N.B. A_{max}: CO_2 saturated assimilation rate. CCP: CO_2 compensation point. K_C^{air}: Modelled Michaelis constant of rubisco for CO_2 in 21% O_2 air. $V_{c\,max}$: Modelled maximum carboxylation rate of rubisco at assay temperature of 30°C. $V_{c\,max}^{25°C}$: Modelled maximum carboxylation rate of rubisco adjusted to 25°C. J_{max}: Maximum electron transport rate at assay temperature of 30°C. $J_{max}^{25°C}$: Maximum electron transport rate adjusted to 25°C. A_C/A_J: Modelled CO_2 concentration at transition between rubisco-limited and ribulose-1,5-bisphosphate (RuBP) regeneration-limited photosynthesis.

Appendix D, Table S5. Stomatal kinetic parameters derived from the temporal profiles of plant gas exchange responses to step changes in light intensity between 100 µmol m^{-2} s^{-1} and 1000 µmol m^{-2} s^{-1}. Values represent mean ± 1 S.E. All data represent an average across three independent single copy lines (n = 3 – 4 per individual line). Differences between transgenic plants and WT are assessed by Fisher LSD post-hoc analysis following a two-way ANOVA, where letters above each box represent statistically significant differences in mean values ($p \leq 0.05$). Significant differences to WT are also highlighted in bold for ease of visualisation.

	WT	**Cyto-*Sm*NOX**	**Stromal-*Sm*NOX**	**CIMS-*Sm*NOX**
λ (minutes)	1.7 ± 0. [a]	1.1 ± 0.1 [a]	1.4 ± 0.2 [a]	1.7 ± 0.3 [a]
k_i (minutes)	9.9 ± 0.7 [a]	8.6 ± 0.7 [a]	8.0 ± 0.5 [a]	10.6 ± 0.8 [a]
SI_{max} (mol m^{-2} s^{-1} min^{-1})	0.065 ± 0.007 [a]	0.101 ± 0.018 [a]	0.096 ± 0.015 [a]	0.046 ± 0.003 [a]
k_d (minutes)	2.0 ± 0.2 [a]	2.5 ± 0.3 [a]	2.1 ± 0.2 [a]	2.1 ± 0.3 [a]

N.B. λ: Initial lag in response time of stomatal conductance upon step increase in light intensity from 100 µmol m^{-2} s^{-1} and 1000 µmol m^{-2} s^{-1}. k_i: Time taken to achieve new steady state stomatal conductance upon step increase in light intensity from 100 µmol m^{-2} s^{-1} and 1000 µmol m^{-2} s^{-1}. SI_{max}: Maximal rate of stomatal opening upon step increase in light intensity from 100 µmol m^{-2} s^{-1} and 1000 µmol m^{-2} s^{-1} min^{-1}. k_d: Time taken to achieve new steady state stomatal conductance upon step decrease in light intensity from 1000 µmol m^{-2} s^{-1} and 100 µmol m^{-2} s^{-1}.

D.5 References

Adrian, J. et al. (2015) 'Transcriptome dynamics of the stomatal lineage: Birth, amplification, and termination of a self-renewing population', Developmental Cell, 33(1), pp. 107–118. Available at: https://doi.org/10.1016/j.devcel.2015.01.025.

Armenteros, J.J.A. et al. (2019) 'Detecting sequence signals in targeting peptides using deep learning', Life Science Alliance, 2(5). Available at: https://doi.org/10.26508/lsa.201900429.

Duursma, R.A. (2015) 'Plantecophys - An R package for analysing and modelling leaf gas exchange data', PLoS ONE, 10(11). Available at: https://doi.org/10.1371/journal.pone.0143346.

Emanuelsson, O. et al. (2000) 'Predicting subcellular localization of proteins based on their N-terminal amino acid sequence', Journal of Molecular Biology, 300(4), pp. 1005–1016. Available at: https://doi.org/10.1006/jmbi.2000.3903.

Engler, C. et al. (2009) 'Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes', PLoS ONE, 4(5). Available at: https://doi.org/10.1371/journal.pone.0005553.

Engler, C. et al. (2014) 'A Golden Gate modular cloning toolbox for plants', ACS Synthetic Biology, 3(11), pp. 839–843. Available at: https://doi.org/10.1021/sb4001504.

Gantner, J. et al. (2018) 'Peripheral infrastructure vectors and an extended set of plant parts for the Modular Cloning system', PLoS ONE, 13(5). Available at: https://doi.org/10.1371/journal.pone.0197185.

Grützner, R. et al. (2021) 'High-efficiency genome editing in plants mediated by a Cas9 gene containing multiple introns', Plant Communications, 2(2). Available at: https://doi.org/10.1016/j.xplc.2020.100135.

Marshall, B. and Biscoe, P. V. (1980) 'A model for C3 leaves describing the dependence of net photosynthesis on irradiance', Journal of Experimental Botany, 31(1), pp. 29–39. Available at: https://doi.org/10.1093/jxb/31.1.29.

McAusland, L. et al. (2016) 'Effects of kinetics of light-induced stomatal responses on photosynthesis and water-use efficiency', The New phytologist, 211(4), pp. 1209–1220. Available at: https://doi.org/10.1111/nph.14000.

Stinziano, J.R. et al. (2020) 'Photosynthesis: tools for plant ecophysiology & modeling', R package version.

Stinziano, J.R. et al. (2021) 'Principles of resilient coding for plant ecophysiologists', AoB PLANTS, 13(5). Available at: https://doi.org/10.1093/aobpla/plab059.

Vialet-Chabrand, S., Dreyer, E. and Brendel, O. (2013) 'Performance of a new dynamic model for predicting diurnal time courses of stomatal conductance at the leaf level', Plant, Cell and Environment, 36(8), pp. 1529–1546. Available at: https://doi.org/10.1111/pce.12086.

Weber, E. et al. (2011) 'A modular cloning system for standardized assembly of multigene constructs', PLoS ONE, 6(2). Available at: https://doi.org/10.1371/journal.pone.0016765.

Woo, H.R. et al. (2016) 'Programming of plant leaf senescence with temporal and inter-organellar coordination of transcriptome in arabidopsis', Plant Physiology, 171(1), pp. 452–467. Available at: https://doi.org/10.1104/pp.15.01929.

'It is just a brief moment in time.'
– Mum