WHAT CAN WE LEARN ABOUT PLANT EVOLUTION FROM A ROBUST PHYLOGENETIC FRAMEWORK?

Tom Carruthers

Department of Plant Sciences

Wolfson College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2019

CONTENTS

ACKNOWLEDGEMENTS

There are several people who I would like to thank for having enabled me to complete this project, and have such an enjoyable and rewarding experience whilst doing so. First, I would like to thank my supervisor Robert Scotland. The hours of discussion with him over the last four years have been extremely stimulating, and acted as a catalyst for the questions and problems I have investigated. He has also taught me, and given me the confidence, to discuss and investigate problems from original perspectives, and be unencumbered by the legacy of other scientists, whilst at the same time providing sceptical and thorough criticism of work that I have presented. From our research group in Oxford, I would also like to thank Pablo Muñoz-Rodíguez for having assembled the immense molecular dataset for *Ipomoea*, and performing much of the original phylogenetic analyses within *Ipomoea*, upon which many of my studies of divergence times and diversifications rates were based. I would also like to thank John Wood, whose taxonomic work for *Ipomoea* has greatly benefitted this project by providing a context within which to evaluate the more purely theoretical questions in this project. I would also like to thank John for accompanying me on a field trip to Bolivia, Paraguay, and Argentina, in 2016. The opportunity to spend a month in the field with one of the world's most prolific plant taxonomists provided me with an invaluable insight into the work that is required beyond computational analyses in order to make inferences about the natural world. From Oxford I would also like to thank colleagues, past and present, with whom I have shared office space and made my time in the Department of Plant Sciences enjoyable. This includes members of Robert Scotland's group – Pablo Muñoz-Rodríguez, Alex Sumadijaya, Tom Wells, Zoë Goodwin – and also members of other groups – Edgar Wong, Gail Stott, Claudia Havrenack, Louise Hill. I would also like to thank Timothy Walker for the great teaching and general encouragement when I was an undergraduate, and also for the encouragement during the interview I had at Somerville in 2011.

From beyond Oxford I would like to thank Mike Sanderson for stimulating discussions about rate variation. The two weeks I spent with him in Arizona were two of the most productive weeks of this entire project, and have underpinned much of the work which was subsequently carried out in Chapter 2. I would also like to thank Tiina Särkinen for her initial guidance in divergence time

estimation, and Colin Hughes for his guidance in assembling literature relating to phylogenetics and the assembly of different biomes.

Beyond those who are directly related to my work, I would like to thank the friends, both in Oxford and beyond, who have supported me throughout the time of my project. I would also like to thank the members of the Thames Valley Leeds United supporters club for being so welcoming, and for so many entertaining days out watching Leeds United around the country. I would also like to thank the homeless population of Oxford. The position you are in is entirely unjust, yet the strength and goodwill that you show is inspiring. From this group I would like to give particular thanks to Adrian and Theodora for your kindness, and many interesting conversations.

Finally, and most importantly, I would like to thank my family. Without the support you have given me for the past 25 years, nothing that I have accomplished would have been possible.

ABSTRACT

A robust molecular phylogeny provides insights into evolutionary history because it illustrates the order that different taxa have diverged from each other, and the changes that have accumulated between different taxa over evolutionary history. However, a molecular phylogeny does not provide information about a range of other fundamental parameters in macroevolutionary research. This includes the absolute time-scale over which a clade has evolved, and the rate that different clades have diversified. In order to estimate absolute time-scales (referred to as divergence time estimation), and infer net diversification rates, speciation rates, and extinction rates (referred to as diversification parameter estimation), additional analyses are therefore required. As robust phylogenies are inferred for an increasing number of groups, and complex analyses are performed to infer macroevolutionary parameters in a greater variety of contexts, evaluating the robustness of methods for inferring divergence times and diversification parameters is more important than ever. Here, I evaluate the robustness of methods for inferring divergence times and diversification parameters. I show that even in the context of genomic scale datasets, among-branch-substitution-rate-variation that acts consistently across entire genomes causes considerable error in divergence time estimates, regardless of the quantity of molecular sequence data that is sampled. I then show that methods that implement multiple fossil calibrations with relaxed clock methods, which are designed to account for among-branch-substitution-rate-variation, make unrealistic assumptions about the fossil record. I show that these unrealistic assumptions are likely to be a further source of error in divergence time estimates. I then characterise the implications of these problems for estimating divergence times and diversification parameters within *Ipomoea*, and show that they can lead to very uncertain inferences. However, I show that if questions are carefully framed, and the assumptions of different methods fully accounted for, robust inferences can be made. This includes the inference that the storage root of the Sweet Potato (*Ipomoea batatas*) evolved in pre-human times, challenging the existing paradigm that it evolved relatively recently as a result of human domestication. It also includes the inference that there is a significant increase in net diversification rates for a clade of Neotropical *Ipomoea* that is of a scale equivalent to some of the most iconic radiations in the plant

kingdom. Taken together, this thesis illustrates fundamental problems that underlie methods in macroevolutionary research, but highlights that when methods are used in the right context, they can serve as a basis for making novel and robust inferences about the natural world.

1. INTRODUCTION

Humans have sought to explain the origins of life for centuries (Linnaeus 1735; Darwin 1859; Nelson and Platnick 1981; Rieppel 1988). Typically, explanations have been sought for the processes that have led to the origin of different taxa and morphological traits, and the processes that have shaped geographical distribution patterns. Although the development of such explanations is a central motivation in modern macroevolutionary research, explanations for these different aspects of the origins of life have not always been discussed in an explicitly evolutionary context.

Regardless, many of the ideas discussed in pre-evolutionary times underpin macroevolutionary research programmes today. This includes the relative importance of dispersal and adaptation, and how, over time, the interaction between these processes has shaped the distribution of extant taxa (Buffon 1761, 1766; Linnaeus 1781; Humboldt 1816; Candolle 1832). More generally, since the time of the ancient Greeks, the idea that there exists a *true* order in nature has played a central role in the development of explanations for the origins of life (Nelson and Platnick 1981; Rieppel 1988).

Nonetheless, an evolutionary perspective, that was most explicitly set out by Darwin (1859), has transformed the way in which the origins of life are explained. This is particularly the case with respect to phylogenetic systematics (Hennig 1966), one of the most important frameworks for conducting macroevolutionary research. Specifically, both the inference and interpretation of phylogenies are heavily dependent on an evolutionary context. First, the inference of a phylogeny requires an understanding and interpretation of character evolution (Camin and Sokal 1965; Wilson 1965; Hennig 1966; Jukes and Cantor 1969; Farris 1977; Farris and Kluge 1979; Tavaré 1986). Subsequently, the inferred phylogeny reflects a “true order”. Interpretation of this in the light of evolution provides a basis for understanding the order that taxa diverged, and the changes that have accumulated between taxa over evolutionary history (Woese and Fox 1977; Nelson and Platnick 1981; Ridley 1983; Lipscomb 1985; Maddison 1990; Pagel 1994, 1999; APG 1998; Soltis et al. 2000; Donoghue et al. 2001; Huelsenbeck et al. 2003; Ree et al. 2005).

Over the last 20 years, molecular sequence data has increasingly been used to construct phylogenies. Although molecular phylogenies were originally constructed with only one or a small number of molecular markers (for example Woese and Fox 1977; Lipscomb 1985; Soltis et al. 2000), they can now be constructed by analysing hundreds of genes or even entire genomes (for example Nevado et al. 2016; Muñoz-Rodríguez et al. 2018, 2019). By enabling more characters to be analysed, the states of which can be easily coded as one of four DNA bases, large molecular datasets have enabled robust phylogenies to be constructed for many groups (Atchison et al. 2016; Nevado et al. 2016; Contreras-Ortiz et al. 2018; Muñoz-Rodríguez et al. 2018, 2019; Nürk et al. 2018). Coupled with major increases in the complexity of methodologies for making inferences about macroevolutionary history from phylogenies (Drummond et al. 2006; Madison et al. 2007; FitzJohn 2012; Condamine et al. 2013; Rabosky 2014; Höhna et al. 2016), this has acted as a basis for increasingly sophisticated macroevolutionary studies (Spriggs et al. 2015; Atchison et al. 2016; Lagomarsino et al. 2016; Nevado et al. 2016; Cardillo et al. 2017; Contreras-Ortiz et al. 2018; Folk et al. 2019; Nürk et al. 2019).

Although a phylogenetic framework may act as a useful basis for making inferences about macroevolutionary history, phylogenetic based macroevolutionary studies have fundamental similarities to much earlier attempts to explain the origins of life (Nelson and Platnick 1981; Rieppel 1988; Sober 1988). Specifically, they rely on observations about the present, and the implementation of a theoretical and methodological framework for making inferences about the past (Sober 1988). As such, two fundamental questions about the validity of macroevolutionary research remain. Do observations about the present contain sufficient information to make the inferences that we aim to make about the past? Is the theoretical and methodological framework within which we work sufficient? When these two questions are considered together, and in the context of modern phylogenetic based macroevolutionary studies, they lead to a single broader question: what can be inferred about macroevolutionary history from a robust phylogenetic framework?

Here, I investigate this question specifically in the context of plant macroevolution. In order to do this, I first discuss recent methodological developments in macroevolutionary research to

provide a basis for understanding the type of information upon which different methods rely, and the directions in which new methods are developing. I then outline recent changes in understanding of plant macroevolution that these methodological developments have led to. This provides a basis for studies in the remainder of this thesis that characterise how and to what extent, in the context of a robust phylogenetic framework and modern methodologies, observations at the present can be used as a basis to make inferences about the past. Taken together, this will enable me to address the question: what can we learn about plant evolution from a robust phylogenetic framework?

RECENT METHODOLOGICAL DEVELOPMENTS IN MACROEVOLUTIONARY RESEARCH

Divergence times

Placing the branching events of a molecular phylogeny on an absolute timescale – referred to as *divergence time estimation* – is one of the most fundamental steps in macroevolutionary research. To estimate divergence times, the lengths of branches in a phylogeny must be transformed such that they are equal to time. This requires additional evidence and assumptions beyond those that are required when inferring a phylogeny. This is because the molecular sequence data from which a phylogeny is inferred only provides direct information about the order that different sequences have diverged, and the total number of substitutions that have occurred along each branch. The total number of substitutions is in turn a product of the temporal duration and substitution rate for each branch (Zuckerkandl and Pauling 1962, 1965; Margoliash 1963).

A critical issue, that has been of fundamental importance throughout the entire history of divergence time estimation, relates to assumptions about the extent to which substitution rates vary between branches. In the simplest case, it can be assumed that the rate is the same on every single branch, such that sequence evolution corresponds to the strict molecular clock model (Zuckerkandl and Pauling 1962, 1965; Margoliash 1963; Miyata 1980; Baldwin and Sanderson 1998). This simple assumption underpins how the first methods of molecular divergence time estimation were developed by Zuckerkandl and Pauling (1962; 1965). Provided that the rate truly is the same on every branch, a

single accurate rate estimate enables the estimation of accurate divergence times throughout the phylogeny.

By contrast, a more complex but likely realistic assumption would be that substitution rates vary among branches (Langley and Fitch 1974; Britten 1984; Gillespie 1989, 1991; Bromham et al. 1996). In this context, relaxed molecular clock models that infer different rates for individual branches can be used, as is the case in the majority of modern studies (Sanderson 1997, 2002; Thorne et al. 1998; Kishino et al. 2001; Drummond et al. 2006; Drummond and Suchard 2011; Heath et al. 2012; Lartillot et al. 2016). Given that the molecular sequence data only provides direct information about the expected number of substitutions along each branch, the implementation of relaxed clock methods requires additional assumptions to produce specific parameter estimates for rates and times on individual branches (Sanderson 1997).

One of the first relaxed clock methods was a penalised likelihood approach developed by Sanderson (1997, 2002). This assumes that rates are inherited between ancestral and descendant branches in a phylogeny, with the degree to which they are inherited being determined by an inferred smoothing value.

A greater variety of relaxed clock methods have been implemented in a Bayesian framework where prior probability distributions are used to make assumptions about parameter values. As with the penalised likelihood approach of Sanderson, Bayesian models can be implemented that assume rates are inherited between ancestral and descendant branches (Thorne et al. 1998; Kishino et al. 2001). Alternatively, uncorrelated models can also be implemented in which the rate for each branch is drawn from a probability distribution, the value for which is entirely independent of the inferred rate for the parent branch (Drummond et al. 2006). Further models have been developed where rates belong to one of a set of discrete rate categories (Drummond and Suchard 2011; Heath et al. 2012), whilst some of the most complex methods combine the attributes of different models in a single analysis in an attempt to capture more accurately the different aspects of rate variation (Lartillot et al. 2016).

In the genomics era, where hundreds of loci can be analysed simultaneously, an emerging point of discussion has been how best to analyse rate variation across large numbers of loci. In this context, rate variation can become increasingly complex. Rates can potentially vary among loci, vary among branches, and simultaneously vary among branches and loci. The implications of these different classes of rate variation for divergence time estimation, and the ways in which such datasets should best be analysed, is unclear (Ho 2014; dos Reis et al. 2014; Zhu et al. 2015). In many cases, a single model of rate variation is implemented that infers the same parameter values for all loci. However, this may mean that important aspects of rate variation are ignored. As such, more complex methods have been developed whereby different parameter values are inferred for individual loci (dos Reis et al. 2014).

In order to estimate absolute rates, and place a phylogeny on an absolute timescale, assumptions about time are also necessary. Typically, these assumptions are made by implementing fossil calibrations. In widely used node calibration methods, the oldest fossil for a clade is used as a minimum constraint for the age of the clade. However, to estimate divergence times, assumptions are also required about how much older a clade is than its oldest fossil. Although necessary, such assumptions can be hard to justify (Donoghue and Benton 2007; Marshall 2008; Ho and Phillips 2009; Heath 2012; Warnock et al. 2012, 2015; Donoghue and Yang 2016). Unlike a minimum constraint, where a clade must be older than its oldest fossil, there is often no direct evidence from which to derive assumptions about how much older a clade is likely to be than its oldest fossil.

In some cases, the assumption of how much older a clade is than its oldest fossil is simply implemented as a maximum constraint, such that the clade age is assumed to lie somewhere between the minimum and maximum (Sanderson 1997; Sanderson and Baldwin 1997; Ho and Phillips 2009; Magallón et al. 2015, Donoghue and Yang 2016). Maximum constraints can be derived from the fossil record by using taphonomic controls, whereby the presence of structurally similar fossils of close relatives to the lineage of interest in a given sediment, but absence of fossils of the lineage of interest from that sediment, can represent evidence that the lineage of interest post-dates those sediments (Benton and Donoghue 2007; Warnock et al. 2015). Such an approach has been used to

derive a maximum constraint for the origin of the eudicots. The oldest eudicot tricolpate pollen grains appear in the fossil record approximately 130 Ma (Doyle et al. 1977). However, the pollen fossil record extends into sediments that are considerably older than this. This can act as evidence that conditions were appropriate for fossilisation of pollen, and that eudicot pollen first appears around 130 Ma because this approximately corresponds to the time that eudicots evolved (Doyle et al. 1977; Bell et al. 2005; Magallón and Sanderson 2005; Smith et al 2010; Magallón et al. 2013, 2015).

When there is a sufficient number of fossils for a given clade, maximum constraints can also be derived by modelling the temporal distribution of fossils within the clade, and using this as a basis to calculate a confidence interval for the age of the clade. Such an approach was developed by Marshall (2008), and implemented in a landmark study that estimated divergence times throughout angiosperms (Magallón et al. 2015) and which has been used as a point of reference in hundreds of subsequent studies.

Aside from the fossil record, other lines of evidence can sometimes be used to justify maximum constraints. For example, if a clade is endemic to an oceanic island or a distinctive habitat, it could be assumed that the clade is no older than that island or habitat (Sanderson and Baldwin 1998; Ho and Phillips 2009).

Although there are several methods for deriving maximum constraints, the implementation of such methods is often problematic. In many clades, especially plants, the fossil record of the relevant group is extremely sparse (Särkinen et al. 2013; Mitchell et al. 2016; Cardillo et al. 2017; Folk et al. 2019). When only a limited number of fossils can be assigned to a clade, it is impossible to meaningfully implement methods that rely on taphonomic controls, or on modelling the temporal distribution of the fossil record. Further, methods based on modelling the fossil record are often implemented such that they make simplistic and unrealistic assumptions, for example, by assuming uniformity in the preservation rate of fossils (Marshall 2008). Problems related to deriving maximum constraints directly from the fossil record are important, because other lines of evidence for deriving maximum constraints, such as the age of a distinctive habitat or island, are not available in many groups.

In addition to maximum constraints, a range of other assumptions can be made about clade ages in relation to oldest fossil ages. In some cases, the fossil calibration is simply implemented as a point calibration such that the age of the oldest fossil is assumed to be equal to the age of the clade (Sanderson 1997; Near et al. 2005; Ho and Phillips 2009). Alternatively, in Bayesian analyses calibration densities are often used whereby a probability distribution is used to describe the probable age of a clade in relation to that of its oldest fossil (Thorne et al. 1998; Drummond et al. 2006; Yang and Rannala 2006; Ho and Phillips 2009; Heath 2012). In a majority of studies, such probability distributions are not explicitly justified (for example Särkinen et al. 2013; Mitchell et al. 2016; Cardillo et al. 2017; Folk et al. 2019).

The implications of temporal constraints become more complex when several fossil calibrations are implemented in a single phylogeny, as is typically the case in recent divergence time analyses (Magallón et al. 2013, 2015; Beaulieu et al. 2015; Barba-Montoya et al. 2018; Morris et al. 2018; Folk et al. 2019). It has been shown that implementing several fossil calibrations in a single phylogeny can reduce error caused by any single calibration in a simple case where divergence times are estimated with a strict molecular clock (Yang and Rannala 2006). However, the assumptions and implications of multiple fossil calibrations are likely to be become increasingly complex when a relaxed molecular clock is used. This is because the fossil calibrations are being used as a framework to infer times and rates for individual branches (Sanderson 1997; Donoghue and Benton 2007; Magallón et al. 2015; Donoghue and Yang 2016). The relationship between the different fossil calibrations is therefore likely to have a profound impact on rate and time estimates in different parts of the phylogeny, and therefore the relative ages of different clades within the phylogeny. In some recent analyses, large numbers of fossil calibrations have been implemented in a single phylogeny with a relaxed clock to infer divergence times (for example Magallón et al. 2015; Barba-Montoya et el. 2018). It has been suggested that this leads to more robust inferences. However, the robustness of such analyses will remain sensitive to the different fossil calibrations that have been implemented, which in this context affect estimates of relative clade ages. Given the often-problematic nature of the

assumptions associated with each fossil calibration (outlined above), this sensitivity represents a potential vulnerability of such analyses.

As an alternative to node calibration methods, and to avoid the arbitrary assumptions outlined above, tip-calibration methods have been advocated in recent years (Stadler 2010; Ronquist et al. 2012; Gavyryushkina et al. 2014; Heath et al. 2014). Based on the analysis of a morphological character matrix, these methods infer explicit relationships between fossils and extant taxa, and combined with the absolute ages of sampled fossils, use this as a framework for inferring divergence times. However, an important assumption of these methods is that the rate of morphological character evolution can be appropriately modelled. If this is not the case, the timescale over which morphological differences accumulate between sampled taxa will be erroneously inferred. Analysis of morphological data in tip-calibration methods is therefore affected by the same fundamental problem that underlies analyses of molecular data in other methods of divergence time estimation – specifically, the rate of character evolution must be inferred. Given the ambiguity associated with many morphological characters (Scotland et al. 2003), modelling rates of morphological evolution (Lewis 2001) is likely to be even more problematic than modelling rates of molecular evolution, for which the associated problems are outlined above.

One tip-calibration method, the fossilised-birth-death-process (FBDP), has been developed that does not require a morphological character matrix with which to infer relationships between fossils and extant taxa (Stadler 2010; Gavyryushkina et al. 2014; Heath et al. 2014). Instead, the analysis can infer divergence time estimates that account for all possible relationships that satisfy a set of constraints reflecting existing knowledge of the phylogenetic placement of fossil calibrations. This method is likely to be particularly relevant for plants because the fragmentary nature of the fossil record means that in many plant clades morphological matrices are not available.

However, an important additional assumption of current implementations of the FBDP for divergence time estimation is that the fossil recovery rate, which describes the rate that fossils are sampled along a branching process, is constant (Stadler 2010; Gavyryushkina et al. 2014; Heath et al. 2014). There is limited evidence to suggest that this is likely to be the case in empirical datasets (Raup

1972; Signor and Lipps 1982; Smith 2001; Smith and McGowan 2005). Although work continues on developing extensions to the FBDP (for example Stadler et al. 2018), no methods for estimating divergence times with the FBDP that enable the fossil recovery rate to vary have been implemented. Further, the prospects of developing such a method are likely to be limited because they would lead fundamental issues with model identifiability. Taken together, although tip-calibration methods avoid some of the assumptions associated with node calibration methods, they make a range of other assumptions that are potentially problematic.

An important further assumption about time, that typically underlies divergence time analyses performed in a Bayesian framework, is that branching events in a time-calibrated phylogeny are controlled by a process in which there are constant rates of speciation and extinction (Drummond et al. 2006; Drummond and Rambaut 2007; Höhna et al. 2016). This additional assumption interacts with assumptions about time that are derived directly from the fossil record. With node calibration methods, the implications of such an interaction are easily observed by the fact that the "effective time-prior", which expresses the prior probability for the age of a node such that the effect of the fossil calibration and branching process are both accounted for, can be markedly different from the specified fossil calibration (Warnock et al. 2011; Warnock et al. 2015). The assumption of constant rates of speciation and extinction is often likely to be unrealistic (for example Baldwin and Sanderson 1998; Hughes and Eastwood 2006) and the implications of this for estimating robust divergence times is unclear.

Diversification parameters

The inference of a time-calibrated phylogeny can enable the inference of the speciation rate (λ) – describing the rate of branching events over time, the extinction rate (μ) – describing the rate that branches terminate before the present, and the net diversification rate ($\lambda - \mu$). The simplest approaches to estimating these parameters – subsequently referred to as *diversification parameters* – assume they remain constant throughout the entire phylogeny (Nee et al. 1994; Nee 2006). However, such an assumption is often likely to be violated. Therefore, assuming constant diversification parameters is likely to cause important macroevolutionary patterns to be overlooked.

Several more complex models for investigating diversification parameters have therefore been developed. For example, episodic diversification models infer different diversification parameters within different time intervals, but assume that within each time interval, the diversification parameters are the same for every branch (Morlon et al. 2011; Stadler 2011a; Höhna 2014, 2015). Alternatively, density dependent models can be used to investigate the relationship between diversification parameters and the diversity of a particular clade (Nee et al. 1992; Rabosky and Lovette 2008). Meanwhile, trait dependent models have been developed that determine whether diversification parameters are associated with particular traits. These models can account for both discrete and continuous traits, and can be extended to also test for associations between diversification parameters and climatic conditions (Maddison et al. 2007; FitzJohn 2012; Condamine et al. 2013; Morlon et al. 2016).

Some of the most complex models for inferring diversification parameters, and those that have attracted the most controversy, are those that explicitly model differences in diversification parameters among different branches (Alfaro et al. 2009; Morlon et al. 2011; Stadler 2011b; Rabosky 2014). Over the past five years, the most widely used method for inferring branch specific diversification parameters has been Bayesian Analysis of Macroevolutionary Mixtures (BAMM)(Rabosky 2014). In BAMM, a time-calibrated phylogeny can diversify under any number of macroevolutionary processes, each of which has distinct diversification parameters.

Recently, BAMM has faced intense criticism (Moore et al. 2016). Specifically, BAMM has been criticised because it only explicitly maps different diversification processes onto sampled and therefore extant branches in a molecular phylogeny. As such, different macroevolutionary processes on extinct branches are ignored, which in turn may lead to biased parameter estimates (Moore et al. 2016). Although the authors of the original critique of BAMM imply that parameter estimates would be improved if the “erroneous” approach of BAMM was modified, a fundamental point remains that any molecular phylogeny necessarily only samples extant lineages. Wherever extinction is present, the inference of diversification parameters may therefore be inherently problematic, an issue that has itself been highlighted by the original producer of BAMM (Rabosky 2010).

A failure to sample extinct lineages represents a failure to sample lineages that existed in the past but no longer exist now. In many cases, it might also be the case that not all extant lineages are sampled. This is likely to be a further source of error and uncertainty when inferring diversification parameters (Stadler 2009; Höhna et al. 2011; Morlon 2011; FitzJohn 2012).

In addition to these problems that essentially relate to sampling, erroneous divergence time estimates will also cause error in diversification parameter estimates. This is because diversification parameters are a description of the temporal distribution of branching events in a phylogeny. Divergence time estimates and diversification parameter estimates are therefore tightly linked. Nonetheless, divergence time estimates are more fundamental, and are essential for diversification parameters to be inferred.

Explaining patterns in the context of ecological and geographical variables

The construction of a robust temporal framework in which diversification parameters may be inferred, can provide a basis for explaining the evolution of a particular group in the context of ecological and geographical variables. Methods in this field have developed rapidly, and in many cases now enable the importance of different variables to be explicitly quantified (Lagomarsino et al. 2016; Cardillo et al. 2017; Folk et al. 2019). However, a detailed analysis of these methods is beyond the scope of this study. Instead, this study focusses predominantly on divergence times and diversification parameters, the inference of which are fundamental to subsequent studies of the role of ecological and geographical variables for explaining macroevolutionary patterns.

HOW HAVE METHODOLOGICAL DEVELOPMENTS CHANGED OUR UNDERSTANDING OF THE ORIGINS OF PLANT DIVERSITY?

The development of methods for making inferences about divergence times and diversification parameters has led to fundamental changes in discussions about the origins of biodiversity. First, it has highlighted the extent to which diversification parameters can undergo marked changes among different clades (Baldwin and Sanderson 1998; Hughes and Eastwood 2006; Givnish et al 2009; Drummond et al. 2012; Nevado et al. 2016). Perhaps the most remarkable case of

this is for the Andean clade of the genus *Lupinus*, for which net diversification rates are 12 times higher than the background rate (Hughes and Eastwood 2006; Drummond et al. 2012; Nevado et al. 2016). As well as highlighting extreme examples such as this, studies of diversification parameters have fuelled much broader areas of research about the extent to which diversification parameters are similar between different clades, and whether extant diversity is "old", or typically belongs to recently and rapidly diversifying clades (Magallón and Sanderson 2001; Hughes et al. 2013; Koenan et al. 2013, 2015; Spriggs et al. 2015).

An increasingly important focus of studies of diversification parameters has also been to determine the processes that explain inferred differences. These studies have led to conclusions about the relationship between inferred diversification parameters, and extrinsic variables such as the extent of geographical isolation, and ecological or environmental variation (Lagomarsino et al. 2016; Cardillo et al. 2017; Folk et al. 2019). For example, Folk et al. (2019) concluded that increased net diversification rates in Saxifragales initially resulted from an increase in the spatial extent of the temperate biome, such that diversification was "niche neutral". Subsequently, as available niches in the temperate biome filled, continued diversification resulted from increased adaptive evolution, such that new species could inhabit novel niche space (Folk et al. 2019). In contrast to these explanations, which focus on the relationship between diversification and the external conditions of a clade, Nevado et al. (2016) characterised intrinsic changes within lineages undergoing rapid evolutionary radiations. Specifically, they showed that elevated net diversification rates were associated with elevated rates of nucleotide evolution at coding loci (Nevado et al. 2016)

Aside from diversification parameters, one of the most important findings that has resulted from the inference of time-calibrated-phylogenies is the apparent prevalence of long-distance dispersal, and not continental drift (as previously assumed), for explaining disjunct distributions between sister taxa (Nelson and Platnick 1981; Tiffney 1985; Donoghue et al. 2001; Lavin et al. 2004). This is inferred because divergence times between sister taxa significantly post-date the geological events previously hypothesised to explain distribution patterns (Lavin et al. 2004). This emphasis on long-distance-dispersal is consistent with an emerging theme over the last 10 years that it

is “easier to move than to evolve”, such that taxa often exhibit niche conservatism and disperse between areas with similar conditions, rather than adapting to new conditions (Donoghue 2008; Crisp 2009).

As a somewhat related point, the inference of time-calibrated phylogenies has seemingly provided fundamental insights into the assembly of the Earth’s major biomes (Hughes et al. 2013). For example, it has been shown that fire-tolerant taxa that are adapted to Savannah conditions evolved simultaneously on both the African and American continent within the last 5 million years (Simon et al. 2009; Simon and Pennington 2012; Maurin et al. 2014; Pennington and Hughes 2014). This coincides with the rapid spread of fire-prone Savannahs dominated by C4 grasses (Beerling and Osborne 2006; Edwards et al. 2010). Further, it has been shown that adaptations to the Savannah biome have evolved several times in individual clades (Simon et al. 2009). This contrasts to the seasonally dry tropical forest biome (SDTF). Clades of SDTF taxa appear to be considerably older and the transition of taxa into the SDTF biome is far less frequent compared to the Savannah biome (Pennington et al. 2004, 2009; Särkinen et al. 2012). In addition to these examples, time-calibrated phylogenies have provided a context for understanding the assembly of many other biomes (Hughes et al. 2013).

There are nonetheless, many aspects of plant evolution that remain poorly understood. Fundamental questions about the age of land plants, and the ages of major clades within land plants remain shrouded in uncertainty (Magallón et al. 2013, 2015; Beaulieu et al. 2015; Barba-Montoya et al. 2018; Hedges et al. 2018; Morris et al. 2018;). Further, even with the apparent progress set out above, the methods upon which these conclusions are based have serious limitations, as set out earlier. The exact implications of these limitations are not necessarily clear. However, some concerning signs have recently emerged whereby fossils have been discovered that are drastically older than previous divergence time estimates (Wilf and Escapa 2015; Wilf et al. 2017). This in turn may mean that conclusions about macroevolutionary history, such as the ages of different biomes, or prevalence of long-distance-dispersal, need to be re-evaluated.

WHAT CAN WE LEARN ABOUT PLANT EVOLUTION FROM A ROBUST PHYLOGENETIC FRAMEWORK?

As was set out previously, methods for inferring divergence times and diversification parameters have rapidly become increasingly complex, yet important limitations remain. Given the fundamental role that inferences from these methods play in subsequent macroevolutionary studies, it is clearly important to evaluate their robustness, and determine whether they are being used to make inferences that are theoretically possible.

In this thesis, I characterise the extent to which precise and accurate inferences of divergence times and diversification parameters can be made from molecular phylogenies. I address this in the context of the genomics era, and in the context of newly developed and often computationally intensive methodologies. First, I evaluate the implications of different classes of substitution rate variation on divergence time estimates in the context of genomic scale multi-locus datasets. I perform these analyses with simulation experiments, but also compare simulated parameter values to an empirical dataset of *Ipomoea*. Subsequently, I evaluate the implications of using multiple fossil calibrations with relaxed clocks to infer divergence times. This is a commonly used method for inferring divergence times when there is among-branch-substitution-rate-variation. Finally, I apply the theoretical principles that are discussed and developed in the preceding two chapters to an empirical dataset for *Ipomoea*. I discuss what can be inferred about the temporal dynamics of evolution within *Ipomoea*. In this chapter, there is a particular emphasis on the temporal dynamics of the evolution of *I. batatas* (the Sweet Potato).

It is crucially important to note that much of the work in this thesis has a strict theoretical focus. Therefore, I use some methods that are not commonly used in macroevolutionary research, and sometimes present results with respect to variables that are not commonly discussed in macroevolutionary research. The purpose of using these different methods and ways of presenting results is to provide the comparative context with which to investigate fundamental theoretical

principles. These theoretical principles are of fundamental importance to modern methods in macroevolutionary research.

2. THE IMPLICATIONS OF LINEAGE-SPECIFIC RATES FOR DIVERGENCE TIME ESTIMATION

ABSTRACT

Substitution rate variation adds considerable complexity to divergence time estimation in molecular phylogenies. Here, I evaluate the impact of *lineage-specific rates* – defined as among-branch-substitution-rate-variation that acts consistently across the entire genome. I compare its impact to *residual rates* – defined as among-branch-substitution-rate-variation that shows a different pattern of rate variation at each sampled locus, and *gene-specific rates* – defined as variation in the average substitution rate across all branches at each sampled locus. I show that lineage-specific rates lead to erroneous divergence time estimates, regardless of how many loci are sampled. Further, I show that stronger lineage-specific rates lead to increasing error. This contrasts to residual rates and gene-specific rates, where sampling more loci significantly reduces error. If divergence times are inferred in a Bayesian framework, I show that error caused by lineage-specific rates significantly reduces the probability that the 95% highest posterior density (HPD) includes the correct value, and leads to sensitivity to the prior. Use of a more complex rate prior – which has recently been proposed to model rate variation more accurately – does not affect these conclusions. Finally, I show that the scale of lineage-specific rates used in the simulation experiments presented here is comparable to that of an empirical dataset for the angiosperm genus *Ipomoea*. Taken together, these findings demonstrate that lineage-specific rates cause error in divergence time estimates, and that this error is not overcome by analysing genomic scale multi-locus datasets.

INTRODUCTION

Since the proposal of the "molecular clock" hypothesis, which made the important assumption that differences between homologous sequences accumulate at a constant rate over time (Zuckerkandl and Pauling 1962, 1965), there has been a continual interest in estimating divergence times in molecular phylogenies (Hori and Osawa 1979; Miyata et al. 1980; Kumar and Hedges 1998; Aris-Brosou and Yang 2003; Kumar 2005). Often, the key assumption of the "molecular clock" is violated and substitution rates differ between branches in a phylogeny (Langley and Fitch 1974; Britten 1984; Gillespie 1989, 1991; Bromham et al. 1996). This can fundamentally compromise divergence time estimates, even when using methods that incorporate rate variation. This is because the number of substitutions along any particular branch – the parameter directly inferred from molecular sequence data – is a product of the substitution rate and the branch's temporal duration (Gillespie 1991; Sanderson 1997, 2002; Thorne et al. 1998; Britton 2005). Without making assumptions about rate variation or divergence times, distinguishing models with different patterns of rate variation or branch duration can therefore become an intractable problem (Sanderson 1997, 2002; Thorne et al. 1998; Kishino et al. 2001).

Previous studies have evaluated the impact of among-branch-substitution-rate-variation on divergence time estimates, and presented new methodologies to account for its effects (Sanderson 1997, 2002; Drummond et al. 2006; Smith and O'Meara 2012; Tamura et al. 2012). These studies have provided a fundamental basis for understanding how the different assumptions about among-branch-substitution-rate-variation that are expressed by different methodologies affect divergence time estimates.

In contrast to these studies focussing on specific methodologies, Britton (2005) provided a more general analysis of the implications of among-branch-substitution-rate-variation for divergence time estimation. Using a mathematical model in the context of a three-taxon tree, Britton (2005) demonstrated that among-branch-substitution-rate-variation leads to erroneous divergence time estimates, regardless of the length of molecular sequence analysed, whether likelihood or Bayesian

inference was used, and even when the correct model of among-branch-substitution-rate-variation was used. Britton (2005) succinctly highlighted that this error results from the fact that molecular sequence data does not provide information about rates for individual branches.

With increasingly large molecular datasets, in which multiple fossil calibrations may also be implemented, sources of uncertainty in divergence time estimates can become increasingly complex. Several studies have attempted to investigate these sources of uncertainty. Yang and Rannala (2006) and dos Reis and Yang (2013) have described how uncertainty results from either limited molecular sequence data or uncertain fossil calibrations. As such, when a large amount of molecular data is sampled, they conclude that uncertainty stems only from fossil calibrations. Although this framework was developed according to the assumptions of the "molecular clock", they subsequently indicated that it is applicable when there is among-branch-substitution-rate-variation, provided a large number of loci are sampled (Zhu et al. 2015). However, when accounting for among-branch-substitution-rate-variation, they assume it shows different patterns at each sampled locus, rather than acting consistently across the entire genome (Zhu et al. 2015). Ho (2014) has stated that these different interpretations of among-branch-substitution-rate-variation will have important effects on divergence time estimates, but no analyses have been undertaken to characterise these effects in different contexts.

Taken together, previous studies have illustrated the important effects of among-branch-substitution-rate-variation on divergence time estimates, and that these effects can become increasingly complex in genomic scale datasets. However, current understanding of the implications of rate variation in genomic scale datasets does not fully take into account many of the more complex ways in which rates can vary in these datasets, and the effects this can have on divergence time estimates.

Here, I perform simulation experiments to evaluate the extent to which among-branch-substitution-rate-variation that acts consistently across entire genomes leads to error in divergence time estimates. I refer to this class of rate variation as *lineage-specific rates* (Fig. 2.1a). The term *lineage effects* is not used because Gillespie (1989) originally used this term to describe the effect of

rate and time on the number of substitutions. I compare the implications of lineage-specific rates to those of *gene-specific rates* – where the average rate across all branches varies at each sampled locus (Fig. 2.1b), and *residual rates* – where rates vary between branches, but the pattern of among-branch-substitution-rate-variation is different at each sampled locus (Fig. 2.1c). I also evaluate the implications when more than one class of rate variation occurs simultaneously (Fig. 2.1d).

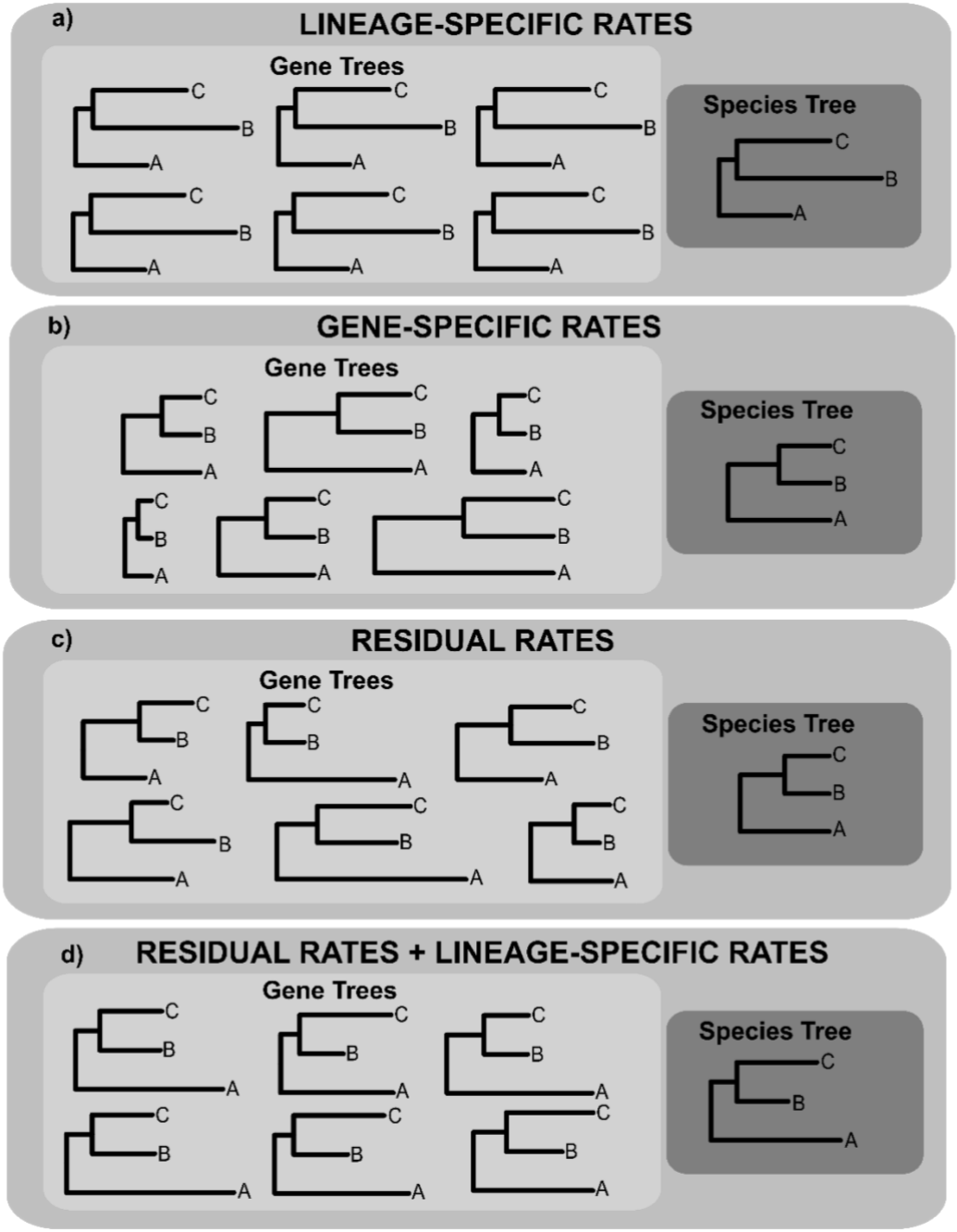

Figure 2.1. (previous page) A summary of different classes of substitution rate variation. The trees shown are phylograms, and in all cases taxa are sampled at the present. In all trees, the two divergence events occurred at time x in the past, and time $2x$ in the past. Therefore, branch length variation reflects substitution rate variation. a) illustrates lineage-specific rates. Among-branch-substitution-rate-variation is observed in each gene tree, but the pattern of among-branch-substitution-rate-variation is the same for each gene tree. The average rate for each branch therefore differs in the species tree. b) illustrates gene-specific rates. The average rate across all branches differs for each gene tree, but does so consistently for all branches. The average rate for each branch is therefore the same in the species tree. c) illustrates residual rates. Among-branch-substitution-rate-variation is observed in all gene trees, but the pattern of among-branch-substitution-rate-variation is different for each gene tree. If a large number of loci are sampled, the average rate for each branch is the same in the species tree. d) illustrates residual rates and lineage-specific rates. Because of the underlying lineage-specific rate, the average rate for each branch differs in the species tree.

I assess how these classes of rate variation influence divergence time estimates when different quantities of molecular data are analysed. In these simulation experiments, the quantity of molecular data is altered by either changing the number of sampled loci, or by changing the sequence length at a single locus. I also compare the performance of two "relaxed clock" priors – one of which has been proposed as particularly effective for divergence time estimation when analysing multi-locus datasets that exhibit rate variation (dos Reis et al. 2014). Finally, I analyse a phylogeny of the angiosperm genus *Ipomoea* to determine whether the lineage-specific rates in the simulation experiments presented here are of a comparable magnitude to those that occur in real datasets.

The issues dealt with in this chapter have significant implications for divergence time estimation because there is widespread evidence that lineage-specific rates are prevalent in many groups, and that they can significantly bias divergence time estimates (Langley and Fitch 1974; Britten 1984; Gillespie 1989, 1991, Bromham et al. 1996; Phillips and Fruciano 2018). Lineage-specific rates have been associated with a range of variables including generation time; efficiency of DNA repair mechanisms; metabolic rates; growth form; or whether the organism is free living or parasitic (Gillespie 1989, 1991; Duff and Nickrent 1997; Lanfear et al. 2013; Ho 2014).

MATERIALS AND METHODS

Summarising the effects of different classes of rate variation

Most simulations were centred on a three-taxon species tree, with a root age of 1.0, and an age for the single internal node of 0.5. The simulations were centred on three-taxon tree to provide the simplest possible context with which to investigate the implications of different classes of rate variation, and where conclusions were not confounded by additional variables inherent in larger trees and more complex analyses (see discussion below on eight-taxon trees). To simulate lineage-specific rates, each branch time duration of this species tree was multiplied by a different lineage-specific rate drawn from a lognormal distribution (mean (μ) = -0.01, standard deviation (σ) = 0.15). This generates a species tree with branch lengths that reflect lineage-specific rates. Four hundred gene trees were then simulated that are identical to this species tree. At this stage, the gene trees are identical to the species tree because gene-specific rates and residual rates have not yet been simulated. To simulate gene-specific rates, all branch lengths in each gene tree were multiplied by a gene-specific rate drawn from a lognormal distribution (μ = -0.01, σ = 0.15). A new gene-specific rate was drawn for each gene tree. To simulate residual rates, each branch length in each gene tree was multiplied by a residual rate drawn from a lognormal distribution (μ = -0.01, σ = 0.15). In these experiments, gene trees were generated with; no lineage-specific rates, gene-specific rates, or residual rates; only lineage-specific rates; only gene-specific rates; only residual rates; and residual rates and lineage-specific rates. Different classes of rate variation were simulated with a custom R script that required the packages phytools (Revell 2012, 2017) and phylobase (Bolker et al. 2017).

DNA sequences of 800 base pairs (bp) were simulated along the branches of each gene tree with a custom R script that used the simSeq function from the package phangorn (Schliep 2011, 2018). Sequences were simulated with a JC model with rate set to 0.05. Simulated sequences were then compiled into one of three datasets. One contained a sequence from a single locus, the second contained sequences from 20 concatenated loci (a total of 16,000 bp), and the third contained sequences from all 400 concatenated loci (a total of 320,000 bp). For each experiment, two further

datasets were also simulated for a single locus of different lengths: one 16,000 bp dataset, and one 320,000 bp dataset.

The divergence time of the single internal node in the three-taxon tree was inferred using RevBayes (Höhna et al. 2016). The topology was fixed to that of the initial species tree, and the root age was fixed at 1.0. The "correct" value for the single unknown divergence time – the internal node – is 0.5. The prior on the branching process was a pure birth (Yule) model. The speciation rate (λ) was sampled from an exponentially distributed prior with a rate parameter of 10.0. Analyses run without sequence data using this prior produced a posterior divergence time that was uniformly distributed between 0.0 and 1.0. Inferred divergence times are therefore unlikely to be influenced by the time prior defined by this branching process. Two different models for the prior on substitution rates were used. A strict clock fixed the rate at 0.05 for all four branches in the three-taxon tree and for all sampled loci. A relaxed clock prior used an uncorrelated lognormal (UCLN) relaxed clock (μ = -3.01, σ = 0.15), in which a separate rate was inferred for each branch in the three-taxon tree, but the same rate was used across all sampled loci. A UCLN relaxed clock was used because this corresponds to the distributions from which rate variation was simulated. The experiment therefore represents a best-case scenario where the "correct" model of rate variation was used to infer divergence times. By contrast, an auto-correlated relaxed clock model was not used because rate variation was not simulated from this model. The use of an auto-correlated model (the "incorrect" model) would likely lead to larger error in divergence time estimates, and provide a more complex context with which to assess the implications of different classes of rate variation. Two hundred replicates of the entire experiment outlined above were performed.

In experiments summarising the effects of different classes of rate variation, the mean rate of the UCLN relaxed clock prior was fixed at 0.05. Although gene-specific rates were simulated from an equivalent distribution, this UCLN relaxed clock may be a poor fit to a dataset in which there are gene-specific rates, and one or a small number of loci has been sampled. In such a case, the mean rate for sampled loci may differ from 0.05. This would contrast to the case where a large number of loci is sampled. In this instance, the mean rate for sampled loci would be far closer to 0.05. Because of this,

an additional experiment was performed where 320,000 bp were sampled from a single locus and gene-specific rates were simulated. In this additional experiment, the mean of the UCLN relaxed clock was not fixed at 0.05, but was instead sampled from a uniform distribution between 10e-6 and 10. The substitution rate prior can therefore correct for the fact that the mean rate for sampled loci is not 0.05. This in turn may lead to more accurate divergence time estimates. This alternative prior was also used in a further experiment where residual rates were simulated, and 320,000 bp were sampled from a single locus. The implications of this alternative prior could therefore be compared in a different context.

Increasing the strength of lineage-specific rates

A further experiment was performed where sequences were simulated with lineage-specific rates of different strengths (from lognormal distributions with σ = 0, 0.15, 0.3 or 0.6). These sequences were also simulated with residual rates, with the same parameters as previously. The unknown divergence time was inferred in RevBayes using two different UCLN relaxed clocks. In one, σ of the UCLN relaxed clock was fixed at 0.15, whilst in the other, σ of the UCLN relaxed clock was altered to exactly match σ of the distribution from which lineage-specific rates had been simulated. When inferring the unknown divergence time, all 400 simulated loci were analysed.

Comparing two different priors for rate heterogeneity

In the above experiments, the UCLN relaxed clock infers a single branch specific rate across all sampled loci. Although this is a widely-used approach, the performance of an alternative prior – the Dirichlet rate prior – which has been suggested for multi-locus datasets that exhibit rate variation, was also evaluated (dos Reis et al. 2014; Zhu et al. 2015).

The Dirichlet rate prior uses a gamma or lognormal distribution for the mean rate amongst all sampled loci ($\bar{\mu}$). With L being the number of sampled loci, a Dirichlet distribution then partitions the total rate ($\bar{\mu} * L$) amongst each of the sampled loci to infer a mean rate for each locus. For each locus, different rates are then inferred for each branch with a UCLN relaxed clock that is specific to that locus and parameterised with the locus specific mean.

The performance of the Dirichlet rate prior was compared to that of the UCLN relaxed clock used previously. The Dirichlet rate prior was parameterised as follows: $\bar{\mu}$ was sampled from a lognormal distribution (μ = -3.01, σ = 0.15), the concentration parameter for the Dirichlet distribution of locus specific rates was 1, and the UCLN model applied to each locus was parameterised with μ = -3.01 and σ = 0.15. The UCLN relaxed clock that inferred a single branch specific rate across all loci was parameterised as previously (μ = -3.01, σ = 0.15). The performance of these two priors was characterised when there were lineage-specific rates, gene-specific rates, and residual rates. These classes of rate variation were simulated as described previously. The unknown divergence time was inferred in RevBayes, using all 400 simulated loci.

Evaluating lineage-specific rates in eight-taxon trees

The implications of lineage-specific rates in eight-taxon trees were also evaluated in order to determine whether findings from experiments with three-taxon trees are likely to be applicable in larger trees. A custom R script that used the function sim.bd.taxa from the package TreeSim was used to simulate trees. Trees were simulated with λ = 1, μ = 0, and the root age was unfixed. Gene trees, rate variation, and DNA sequences were simulated according the same principles as with the three-taxon trees. DNA sequences were simulated with either no rate variation, or lineage-specific rates from a lognormal distribution with μ = -3.01 and σ = 0.15.

Divergence times were inferred using RevBayes. The topology was fixed to that of the initial species tree and the root age was fixed to the correct value. The prior for the branching process was a pure birth (Yule) model with λ fixed to 1 (the correct value). Unlike with the three-taxon trees, analyses run without molecular sequence data produced posterior divergence time estimates that were not uniformly distributed between 0.0 and the root age. Therefore, in this analysis, inferred divergence times are likely to be influenced by the branching process. This analysis therefore provides a less direct illustration of the effects of rate variation, but allows its effects to be investigated in a more complex tree. In the simulated eight-taxon tree λ is constant, and the correct constant λ is implemented with the prior for the branching process. In the experiments presented here this may lead

to reduced error in divergence time estimates. However, it would be misleading to suggest this reduction in error would be relevant to empirical analyses, because in such analyses λ is neither known nor likely to be constant. Two different models for the prior on substitution rates were used. A strict clock fixed the rate at 0.05 for all branches and for all sampled loci. A relaxed clock prior used an uncorrelated lognormal (UCLN) relaxed clock ($\mu = -3.01$, $\sigma = 0.15$), in which a separate rate was inferred for each branch, but the same rate was used across all sampled loci.

An empirical study: quantifying the magnitude of lineage-specific rates in Ipomoea

The magnitude of lineage-specific rates in a phylogeny for *Ipomoea* that was inferred from a dataset of 434 concatenated nuclear genes (Muñoz-Rodríguez et al. 2019 in press) was quantified. All 76 pairs of sister species (terminal taxa) were extracted from this phylogeny. Because these sister pairs have the same branch time duration, by definition, they can provide some insight into the extent of lineage-specific rates. The extent to which lineage-specific rates simulated in experiments presented here are comparable to those that occur in a biological dataset can therefore be evaluated.

Custom R, Revbayes, and Python scripts that were developed for all simulation experiments, and analyses of lineage-specific rates in *Ipomoea*, are available in Electronic Appendix 1.

RESULTS

Results from simulation experiments are presented with respect to the mean posterior estimate (MPE) and the 95% highest posterior density (95% HPD). The MPE illustrates the extent to which the analysis is converging on the "correct" value and extent to which the posterior distribution shifts in different contexts. By contrast, analysis of the 95% HPD provides insights into the width of the posterior distribution, and the probability that the 95% HPD contains the correct value in different contexts. Some researchers may be surprised by the focus that is placed on the MPE, given that many studies of divergence time estimation pay little attention to the MPE. However, this is a theoretical study, and analysis of the MPE is a useful tool for understanding how different classes of rate variation effect divergence time estimates.

Summarising the effects of different classes of rate variation

In the results presented in detail here, the UCLN relaxed clock was used to infer the unknown divergence time. A comparison with the results obtained when using the strict clock follows.

No rate variation

Increasing the number of sampled loci led to reduced error in mean posterior age estimates (Fig. 2.2a). The root mean squared error (RMSE) – which quantifies the magnitude of error in the same units as the simulation experiment – fell from 0.0791 when sampling 1 locus, to 0.0108 when sampling 400 loci (Table 2.1). Increasing the number of sampled loci also caused the mean 95% highest posterior density (HPD) width to decrease from 0.346 to 0.192, and the percentage of replicate experiments that included the correct value in the 95% HPD to increase from 97.5% to 100% (Table 2.1). Increasing the sampled sequence length at a single locus had an indistinguishable effect from increasing the number of sampled loci (Fig. 2.2, Table 2.1).

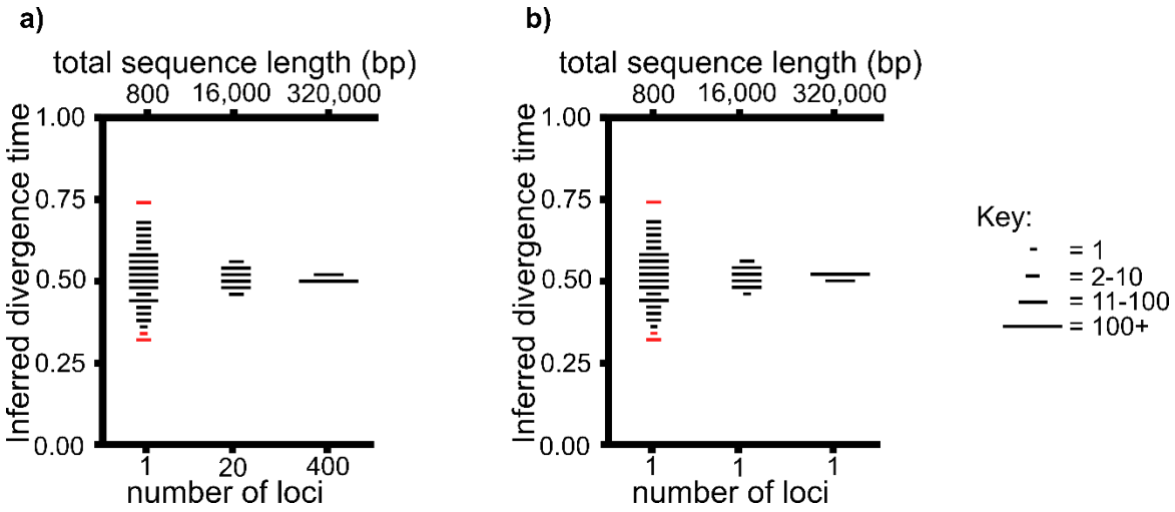

Figure 2.2. The impact of increasing the quantity of sampled molecular sequence data on divergence time estimates when there is no rate variation. In a) the total sampled sequence length increases from 800 to 320,000 bp by increasing the number of sampled loci from 1 to 400 (each locus has a length of 800 bp). In b) the total sampled sequence length increases from 800 to 320,000 bp by increasing the sampled sequence length at a single locus. The unknown divergence time is inferred with a UCLN relaxed clock. Mean posterior age estimates from each experiment are plotted within bins of 0.2. The length of each plotted line corresponds to the number of estimates within each bin following 200 replicate experiments (see key). The line is plotted in red if for more than 50% of replicate experiments plotted within a bin, the 95% HPD did not include the correct value. Otherwise the line is plotted in black.

Gene-specific rates

Increasing the number of sampled loci led to reduced error in mean posterior age estimates (Fig. 2.3a). The RMSE fell from 0.0818 when sampling 1 locus to 0.0114 when sampling 400 loci (Table 2.1). The mean 95% HPD width, and percentage of replicate experiments that included the correct value within the 95% HDP, were similar to when there was no rate variation (Table 2.1).

When the sampled sequence length at a single locus was increased, the reduction in error was considerably less (Fig. 2.3b). The RMSE fell from 0.0818 when sampling 800 bp to 0.0457 when sampling 320,000 bp (Table 2.1). The mean 95% HPD widths were similar to when incrementally more loci were sampled (Table 2.1), whilst fewer replicate experiments included the correct value within the 95% HPD (Table 2.1). When a prior that more explicitly accounts for gene-specific rates was used, and 320,000 bp were sampled from a single locus, the RMSE fell such that it more closely resembled the case where 400 loci were sampled (Table 2.2).

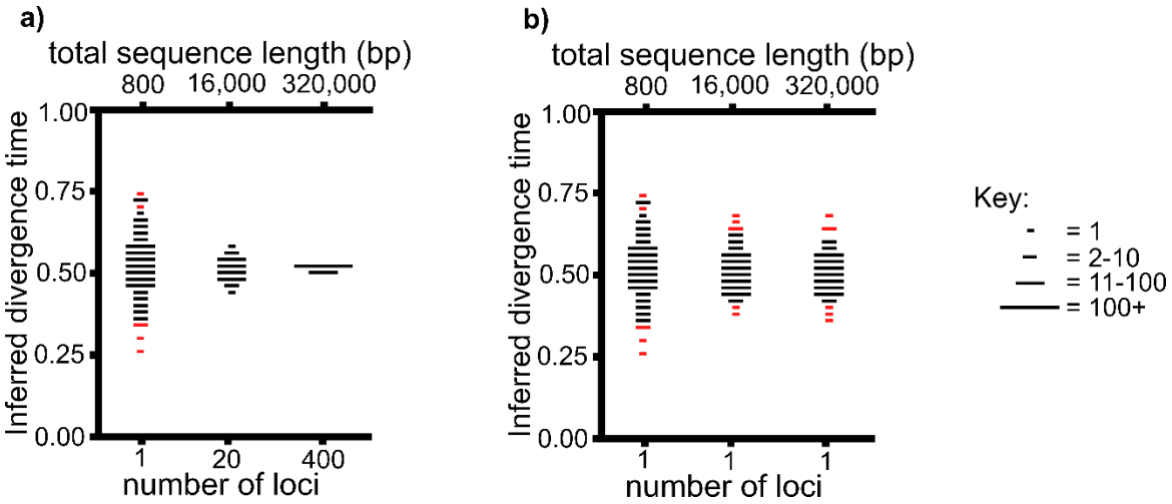

Figure 2.3. The impact of increasing the quantity of sampled molecular sequence data on divergence time estimates when there are gene-specific rates. In a) the total sampled sequence length increases from 800 to 320,000 bp by increasing the number of sampled loci from 1 to 400 (each locus has a length of 800 bp). In b) the total sampled sequence length increases from 800 to 320,000 bp by increasing the sampled sequence length at a single locus. The unknown divergence time is inferred with a UCLN relaxed clock. Points are plotted according to the same bins, sizes and colours as Figure 2.2.

Residual rates

Increasing the number of sampled loci led to reduced error in mean posterior age estimates (Fig. 2.4a). The RMSE fell from 0.0886 when sampling 1 locus, to 0.0106 when sampling 400 loci (Table 2.1). The mean 95% HPD widths, and percentage of replicate experiments that included the correct value within the 95% HPD, were similar to previous experiments in which the number of sampled loci was incrementally increased (Table 2.1).

When the sequence length at a single locus was increased, the reduction in error was considerably less (Fig. 2.4b). The RMSE fell from 0.0886 when sampling 800 bp to 0.0486 when sampling 320,000 bp (Table 2.1). The mean 95% HPD widths were similar to when the number of sampled loci was incrementally increased, whilst fewer replicate experiments included the correct value within the 95% HPD (Table 2.1). When a prior that more explicitly accounts for gene-specific rates was used, and 320,000 bp were sampled from a single locus, the RMSE did not fall compared to when the standard prior was used (Table 2.2).

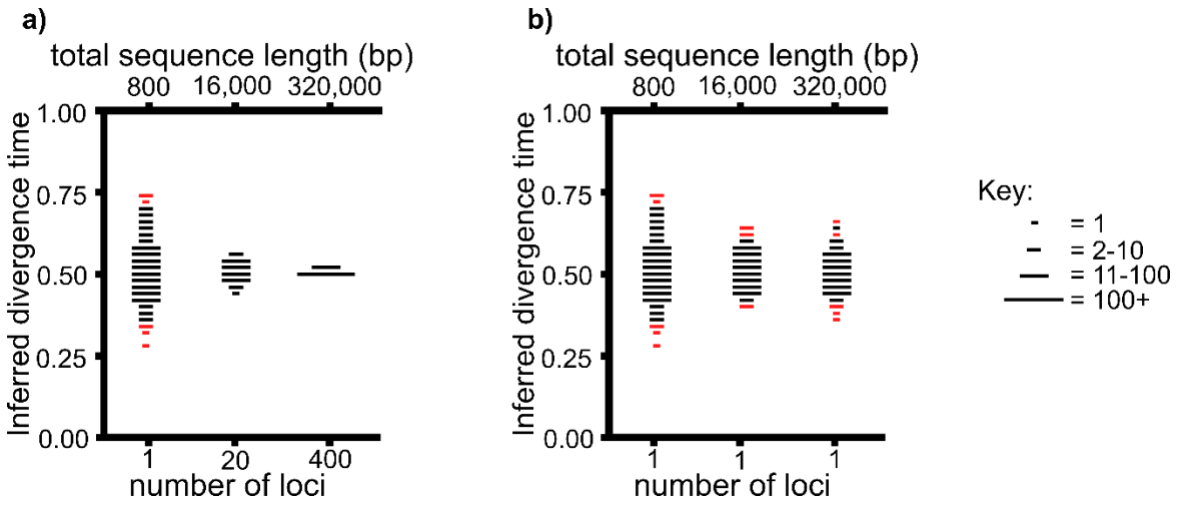

Figure 2.4. The impact of increasing the quantity of sampled molecular sequence data on divergence time estimates when there are residual rates. In a) the total sampled sequence length increases from 800 to 320,000 bp by increasing the number of sampled loci from 1 to 400 (each locus has a length of 800 bp). In b) the total sampled sequence length increases from 800 to 320,000 bp by increasing the sampled sequence length at a single locus. The unknown divergence time is inferred with a UCLN relaxed clock. Points are plotted according to the same bins, sizes and colours as Figure 2.2.

Lineage-specific rates

Increasing the number of sampled loci had a limited impact with respect to reducing error in mean posterior age estimates (Fig. 2.5a). The RMSE fell from 0.0875 when sampling 1 locus, to 0.0510 when sampling 400 loci (Table 2.1). The mean 95% HPD widths were similar to previous experiments (Table 2.1), whilst the percentage of replicate experiments that included the correct value within the 95% HPD was lower than previous experiments in which the number of sampled loci was incrementally increased. Increasing the sampled sequence length at a single locus had a very similar impact to sampling more loci (Fig. 2.5b, Table 2.1).

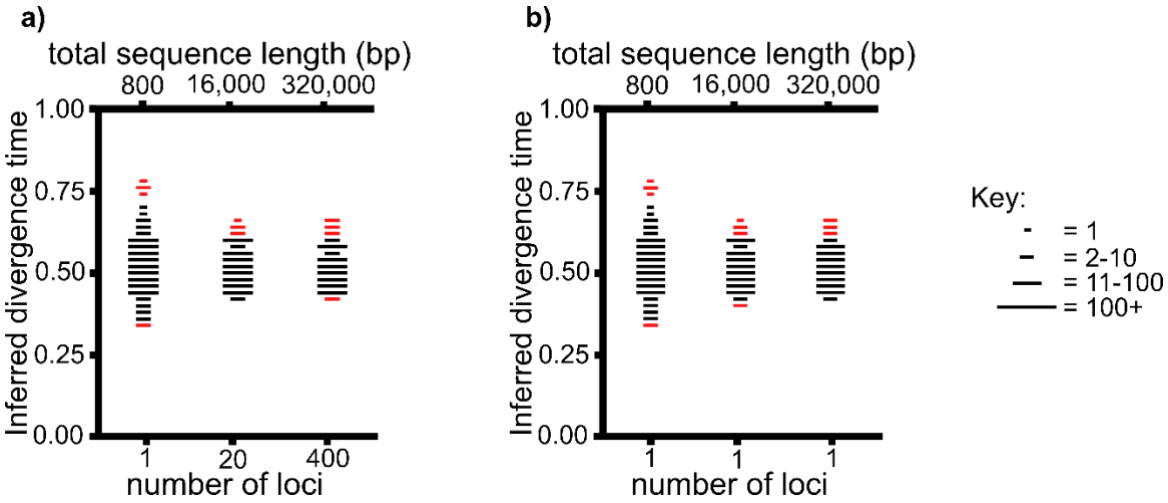

Figure 2.5. The impact of increasing the quantity of sampled molecular sequence data on divergence time estimates when there are lineage-specific rates. In a) the total sampled sequence length increases from 800 to 320,000 bp by increasing the number of sampled loci from 1 to 400 (each locus has a length of 800 bp). In b) the total sampled sequence length increases from 800 to 320,000 bp by increasing the sampled sequence length at a single locus. The unknown divergence time is inferred with a UCLN relaxed clock. Points are plotted according to the same bins, sizes and colours as Figure 2.2.

Lineage-specific rates and residual rates

Increasing the number of sampled loci had a moderate impact with respect to reducing error in mean posterior age estimates (Fig. 2.6a). The RMSE fell from 0.108 when sampling 1 locus, to 0.0518 when sampling 400 loci (Table 2.1). Mean 95% HPD widths were similar to previous experiments, whilst the percentage of replicate experiments that included the correct value within the

95% HDP increased from 87.5% when sampling 1 locus to 95.5% when sampling 400 loci (Table 2.1).

Increasing the sampled sequence length at a single locus had a more limited impact with respect to reducing error in mean posterior age estimates (Fig. 2.6b). The RMSE fell from 0.108 when sampling 800 bp to 0.0743 when sampling 320,000 bp. Mean 95% HPD widths were similar to previous experiments. The percentage of replicate experiments that included the correct value within the 95% HPD fell from 87% when sampling 800 bp to 80.5% when sampling 320,000 bp (Table 2.1).

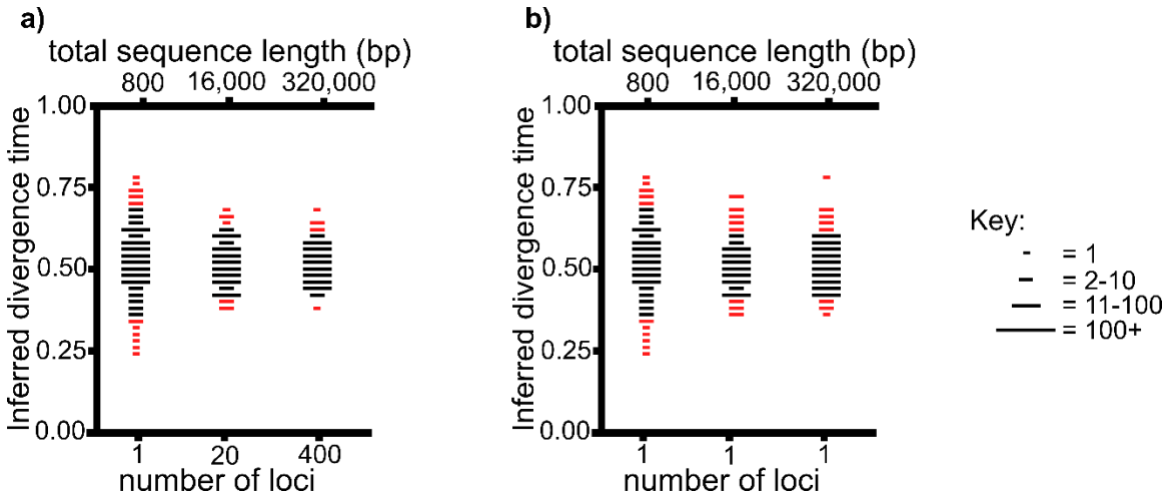

Figure 2.6. The impact of increasing the quantity of sampled molecular sequence data on divergence time estimates when there are lineage-specific rates and residual rates. In a) the total sampled sequence length increases from 800 to 320,000 bp by increasing the number of sampled loci from 1 to 400 (each locus has a length of 800 bp). In b) the total sampled sequence length increases from 800 to 320,000 bp by increasing the sampled sequence length at a single locus. The unknown divergence time is inferred with a UCLN relaxed clock. Points are plotted according to the same bins, sizes and colours as Figure 2.2.

Using a strict clock to infer the unknown divergence time

The distribution of mean posterior age estimates was broadly the same when the unknown divergence time was inferred with the strict clock compared to the UCLN relaxed clock (Table 2.3). However, the mean 95% HPD widths were far narrower. As such, the correct value was included within the 95% HPD in far fewer replicate experiments. This is most notably the case when lineage-specific rates were present and the largest molecular datasets were sampled (either 400 loci, or

320,000 bp at a single locus). In these instances, the 95% HPD included the correct value only about 10% of the time (Table 2.3).

Table 2.1. Comparing the impact of different classes of rate variation when sampling different quantities of sequence and the unknown divergence time is inferred with a UCLN relaxed clock

		Number of loci			Sequence length (bp) at one locus		
		1	**20**	**400**	**400**	**16,000**	**320,000**
No rate variation	**Root mean squared error**	0.0791	0.0208	0.0108	0.0791	0.0199	0.0120
	Mean 95% HPD width	0.346	0.201	0.192	0.346	0.201	0.192
	% of replicates including correct value within 95% HPD	97.5	100	100	97.5	100	100
Gene-specific rates	**Root mean squared error**	0.0818	0.0213	0.0114	0.0818	0.0509	0.0457
	Mean 95% HPD width	0.350	0.202	0.191	0.350	0.203	0.191
	% of replicates including correct value within 95% HPD	96	100	100	96	95.5	96
Residual rates	**Root mean squared error**	0.0886	0.0231	0.0106	0.0886	0.0536	0.0486
	Mean 95% HPD width	0.346	0.202	0.192	0.346	0.203	0.190
	% including correct value within 95% HPD	95	100	100	95	94	95.5
Lineage-specific rates	**Root mean squared error**	0.0875	0.0528	0.0510	0.0875	0.0543	0.0512
	Mean 95% HPD width	0.346	0.203	0.192	0.346	0.202	0.192
	% of replicates including correct value within 95% HPD	95	94	93.5	95	93	95.55
Lineage-specific rates and residual rates	**Root mean squared error**	0.108	0.0554	0.0518	0.108	0.0731	0.0743
	Mean 95% HPD width	0.349	0.203	0.192	0.349	0.202	0.193
	% of replicates including correct value within 95% HPD	87	94	94.5	87	82.5	80.5

Table 2.2. Using a UCLN relaxed clock with a mean sampled from between 10e-6 and 10, 320,000 bp at a single locus is sampled, and there are gene-specific rates or residual rates

Gene-specific rates	**Root mean squared error**	0.00778
	Mean 95% HPD width	0.219
	% of replicates including correct value within 95% HPD	100
Residual rates	**Mean squared error**	0.0570
	Mean 95% HPD width	0.219
	% of replicates including correct value within 95% HPD	94

Table 2.3. Comparing the impact of different classes of rate variation when sampling different quantities of sequence and the unknown divergence time is inferred with a strict clock

		Number of loci			**Sequence length (bp) at one locus**		
		1	**20**	**400**	**400**	**16,000**	**320,000**
No rate variation	**Root mean squared error**	0.0772	0.0183	0.00363	0.0772	0.0171	0.00407
	Mean 95% HPD width	0.288	0.0643	0.0145	0.288	0.0646	0.0145
	% of replicates including correct value within 95% HPD	94.5	93	94	94	94.5	92
Gene-specific rates	**Root mean squared error**	0.0779	0.0185	0.00433	0.0779	0.0375	0.0328
	Mean 95% HPD width	0.288	0.0645	0.0146	0.288	0.0647	0.0145
	% of replicates including correct value within 95% HPD	92	93	91	92	61.5	14.5
Residual rates	**Root mean squared error**	0.0877	0.0199	0.00444	0.0877	0.0527	0.0488
	Mean 95% HPD width	0.287	0.0647	0.0145	0.287	0.0652	0.0144
	% of replicates including correct value within 95% HPD	89.5	91	88	89.5	43	9.5
Lineage-specific rates	**Root mean squared error**	0.0868	0.0527	0.0511	0.0868	0.0542	0.0514
	Mean 95% HPD width	0.288	0.0650	0.0145	0.288	0.0648	0.0146
	% of replicates including correct value within 95% HPD	91	50.5	11.5	91	47	13

Lineage-specific rates + Residual-rates	**Root mean squared error**	0.107	0.0546	0.0504	0.107	0.0717	0.0736
	Mean 95% HPD width	0.290	0.0649	0.0145	0.290	0.0649	0.0146
	% of replicates including correct value within 95% HPD	80.5	51	13.5	80.5	37	10

Stronger lineage-specific rates lead to larger error in divergence time estimates

Stronger lineage-specific rates caused mean posterior age estimates to differ more from the correct value. This was the case regardless of whether σ of the UCLN relaxed clock was corrected such that it was equal to σ of the distribution from which lineage-specific rates were simulated (Fig. 2.8a and b). When σ of the UCLN relaxed clock was corrected, the RMSE increased from 0.00363 when there were no lineage-specific rates, to 0.150 with the strongest lineage-specific rates (Table 2.4). When σ was fixed at 0.15, the RMSE increased from 0.0108 to 0.195 (Table 2.4).

When σ was corrected, the mean 95% HPD width increased from 0.0145 with no lineage-specific rates, to 0.588 with the strongest lineage-specific rates. The percentage of replicate experiments that included the correct value within the 95% HPD showed only a small fall from 94% with no lineage-specific rates, to 92.5% with the strongest lineage-specific rates (Table 2.4). When σ was fixed at 0.15, the mean 95% HPD width initially remained constant at around 0.192, but then showed a slight fall to around 0.176 with the strongest lineage-specific rates (Table 2.4). The percentage of replicate experiments that included the correct value within the 95% HPD decreased from 100% with no lineage-specific rates to 32% with the strongest lineage-specific rates (Table 2.4).

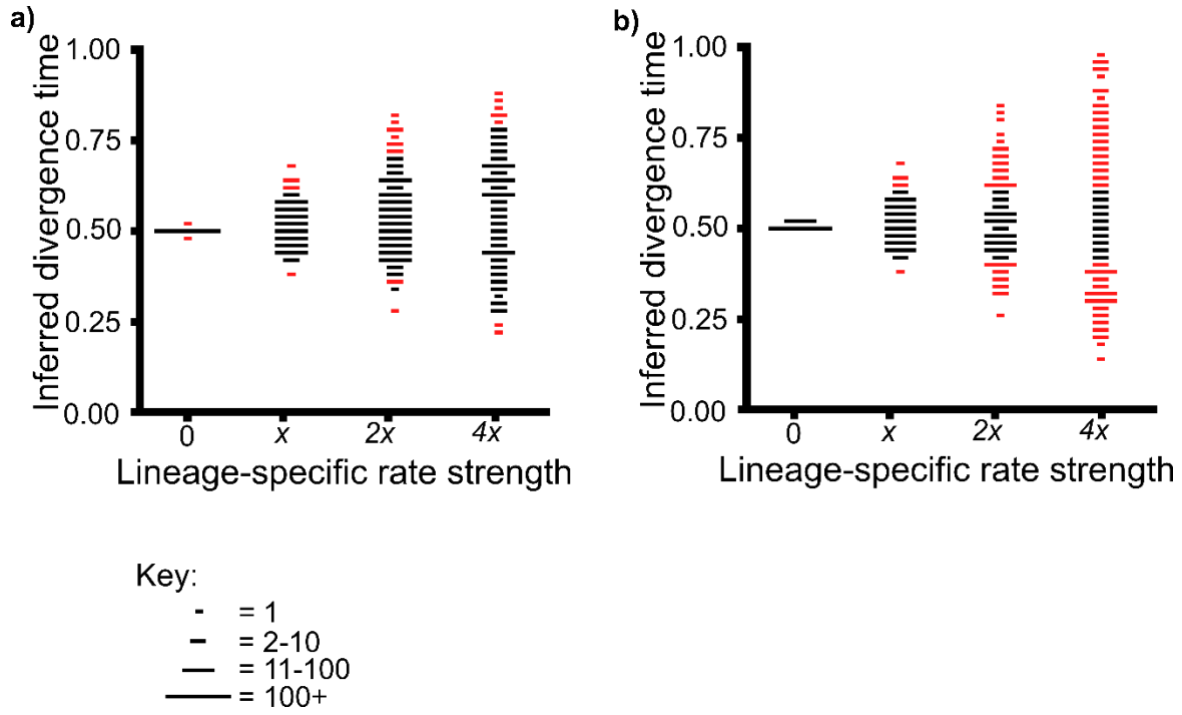

Figure 2.8. The impact of increasingly strong lineage-specific rates on divergence time estimates. Points are plotted according to the same bins, sizes and colours as Figure 2.2. In all cases, the unknown divergence time is inferred from 400 loci. a) σ of the UCLN relaxed clock is corrected such that it exactly matches σ of the distribution from which lineage-specific rates are simulated. b) σ of the UCLN relaxed clock is fixed at 0.15.

Table 2.4. Summarising the effect of increasingly strong lineage-specific rates when 400 loci are sampled

		Lineage-specific rate strength			
		0	***x***	**2*x***	**4*x***
UCLN relaxed clock σ corrected	**Root mean squared error**	0.00363	0.0512	0.108	0.150
	Mean 95% HPD width	0.0145	0.192	0.381	0.588
	% of replicates including correct value within 95% HPD	94	94.5	93.5	92.5
UCLN relaxed clock σ fixed	**Root mean squared error**	0.0108	0.0518	0.106	0.195
	Mean 95% HPD width	0.192	0.192	0.193	0.176
	% of replicates including correct value within 95% HPD	100	94.5	61	32

The Dirichlet rate prior does not reduce error when there are lineage-specific rates

The Dirichlet rate prior did not reduce error in mean posterior age estimates. The RMSE for the Dirichlet rate prior was 0.0599, whilst the RMSE for the UCLN relaxed clock was 0.0552 (Table 2.5). The Dirichlet rate prior did result in considerably narrower 95% HPD widths. For the Dirichlet rate prior, the mean 95% HPD width was 0.0186, whilst for the UCLN relaxed clock, the mean 95% HDP width was 0.190. The Dirichlet rate prior only included the correct value within the 95% HPD in 10% of replicate experiments, whilst the UCLN relaxed clock included the correct value within the 95% HPD in 89% of replicate experiments.

Table 2.5. Comparing the UCLN prior and Dirichlet rate prior when there are lineage-specific rates and 400 loci are sampled

	UCLN Prior	**Dirichlet Rate Prior**
Root mean squared error	0.0552	0.0599
Mean 95% HPD width	0.190	0.0186
% of replicates including correct value within 95% HPD	89.0	10.0

Error resulting from lineage-specific rates is not specific to three-taxon trees

To determine the generality of results in three-taxon trees, a further set of analyses were performed with eight-taxon trees. In this section, the correct divergence time is not fixed at 0.5. Therefore, MPEs are evaluated with respect to the percentage error relative to the correct value, and 95% HPDs are evaluated with respect to their width relative to the MPE.

Lineage-specific rates led to erroneous divergence time estimates in eight-taxon trees. When a UCLN relaxed clock was used and 400 loci were sampled, the mean error for the MPE was 7.785% when there were lineage-specific rates, and was 1.421% when there were no lineage-specific rates (Table 2.6). Mean 95% HPD widths were 29.510% of the MPE when there were lineage-specific rates, and 29.006% of the MPE when there were no lineage-specific rates. The percentage of nodes for which the 95% HPD contained the correct value was 88.9% when there were lineage-specific rates, and 100% when there were no lineage-specific rates. When a strict clock was used, the mean

error for the MPE was similar to when a UCLN relaxed clock was used, especially when there were lineage-specific rates. However, the 95% HPDs were far narrower and the percentage of nodes for which the 95% HPD contained the correct value was lower, and was far lower when there were lineage-specific rates (Table 2.6).

Table 2.6. Assessing the impact of lineage-specific rates in eight taxon trees, when either a strict clock or a UCLN relaxed clock is used to infer divergence times

		Strict clock	UCLN Relaxed Clock
No rate variation	**% error**	0.659	1.421
	Mean 95% HPD width (as % of MPE)	3.132	29.006
	% of nodes for which 95% HPD includes correct value	91.1	100
Lineage-specific rates	**% error**	8.248	7.785
	Mean 95% HPD width (as % of MPE)	3.159	29.510
	% of nodes for which 95% HPD includes correct value	7.8	88.9

Lineage-specific rates in Ipomoea are of a comparable scale to those used in the simulation experiments

Across all sister species comparisons in the *Ipomoea* species tree, there is a mean rate difference of 1.23-fold. Given that the species tree is inferred from 434 concatenated loci, these rate differences are likely to represent lineage-specific rates. This implies stronger lineage-specific rates than in the majority of simulation experiments presented here where rates are drawn from a lognormal distribution with $\sigma = 0.15$. Such a distribution results in a mean rate difference between sister species of 1.19-fold.

DISCUSSION

Lineage-specific rates cause divergence time estimation error

Lineage-specific rates lead to error in divergence time estimates, and increasing the number of sampled loci, or sampled sequence length at a single locus, has a limited effect in reducing this error (Fig. 2.5 and 2.6, and Table 2.1 and 2.3). For example, when lineage-specific rates were present,

increasing the number of sampled loci from 1 to 400 only caused a decrease in error from 17.5% to 10.2% (percentage errors represent the RMSE of mean posterior age estimates as a percentage of the correct value). By contrast, with residual rates, increasing the number of sampled loci from 1 to 400 caused error to decrease from 17.7% to 2.21%. Moreover, increasingly strong lineage-specific rates lead to increasingly large errors in age estimates (Fig. 2.8 and Table 2.4). These results are not specific to the case of the three-taxon tree that was used in most experiments, because lineage-specific rates led to a similar level of error in additional experiments with eight-taxon trees (Table 2.6). The slightly reduced error in the eight-taxon trees is likely to be caused by the tree prior – when divergence times were inferred in experiments with eight-taxon trees the tree prior was correctly parameterised such that it was consistent with the parameters under which the trees were simulated. This is unlikely to be the case in an empirical dataset. A further important result is that the error caused by lineage-specific rates is not reduced by using a more complex rate prior that has been specifically developed for multi-locus datasets. Instead, this more complex prior led to misleadingly precise divergence time estimates (Table 2.5).

Comparing different classes of rate variation reveals why lineage-specific rates are problematic

Increasing the quantity of molecular sequence data enables the relative number of substitutions along each branch to be inferred with less error (Britton 2005). With no rate variation, such that the substitution rate is the same on every branch, this means that the unknown divergence time is inferred with less error (Fig. 2.2, Table 2.1 and 2.3). Further, with no rate variation, and therefore no rate differences between loci, the effect of increasing the sequence length at a single locus is effectively the same as increasing the number of sampled loci (Fig. 2.2, Table 2.1 and 2.3).

With locus specific classes of rate variation (gene-specific rates and residual rates), increasing the number of sampled loci means that the average substitution rate of sampled loci more accurately reflects that of the entire genome. When there are gene-specific rates, this means that unusually fast or slow loci are less likely to cause poor model parameterisation and error in divergence time estimates (Fig. 2.3a, Table 2.1, and 2.3). When there are residual rates, this means that the problem reduces to a "strict" molecular clock because there are no overall rate differences among branches across the entire

genome (Fig. 2.1) – this in turn leads to less error in inferred divergence times (Fig. 2.4a, Table 2.1, Table 2.3). Alternatively, when only the sequence length at a single locus is increased, an average rate across the entire genome cannot be accurately inferred. As such, for both gene-specific rates and residual rates, the reduction in error is far more limited (Fig. 2.3b and 2.4b, Table 2.1, 2.2, and 2.3)

For lineage-specific rates, however, sampling more loci does not lead to the same improvements in accuracy as are possible for residual rates. In this case, the average substitution rate across all sampled loci is different for each branch, regardless of the number of loci that are sampled (Fig. 2.1d). Because of this, sampling more loci has a far more limited effect with respect to reducing error in divergence time estimates, and is no more effective than increasing the sampled sequence length at a single locus (Fig. 2.5, Table 2.1 and 2.3).

When there are residual rates and lineage-specific rates, sampling more loci is marginally more effective at reducing error than increasing the sequence length at a single locus (Fig. 2.6, Table 2.1 and 2.3). This may be because increasing the number of sampled loci reduces error associated with residual rates, whilst error resulting from lineage-specific rates remains.

Interpreting error in the context of 95% HPD intervals

In the discussion of error presented here, I have so far focussed on mean posterior age estimates. This provides a useful framework to evaluate the effect of different classes of rate variation and different quantities of molecular sequence data. However, when inferred in a Bayesian framework, divergence time estimates are often discussed with respect to the 95% HPD.

In this study, the 95% HPD width is sensitive to the quantity of data that is analysed (increasing the number of sampled loci and the sequence length at a single locus have an identical effect), and the rate prior that is used. The class of rate variation with which sequences were simulated did not affect the 95% HPD width – except potentially for very strong lineage-specific rates (Table 2.1, 2.3, and 2.4).

In initial experiments that evaluated different classes of rate variation, and where the unknown divergence time was inferred with a UCLN relaxed clock, the variance of the UCLN relaxed

clock was equal to the variance of the distributions from which rate variation was simulated. Thus, when a single class of rate variation was simulated, a high percentage (approximately 95% or more) of replicate experiments included the correct value within the 95% HPD (Fig. 2.2, 2.3, 2.4, 2.5, and Table 2.1).

However, when there was more than one class of rate variation (lineage-specific rates and residual rates), and 320,000 bp from a single locus were sampled, only 80.5% of replicate experiments included the correct value within the 95% HPD (Fig. 2.6b and Table 2.1). In this case, the variance of the UCLN relaxed clock appears to have been insufficient to account for both classes of simulated rate variation. This contrasts to the case where 400 loci were sampled, and lineage-specific rates and residual rates were present. In this case, 94.5% of replicate experiments included the correct value within the 95% HPD. Here, sampling more loci is likely to have reduced the impact of residual rates. As such, the variance of the UCLN relaxed clock was sufficient to account for remaining rate variation that stemmed from lineage-specific rates (Fig. 2.6a and Table 2.1).

Two further cases highlight that with lineage-specific rates, the utility of the 95% HPD is highly sensitive to the variance of the rate prior, even when the maximum number of loci is sampled. First, when strong lineage-specific rates were simulated, and the variance of the UCLN relaxed clock was fixed at a low value, only 32% of replicate experiments included the correct value within the 95% HPD (Fig. 2.8a, Table 2.4). Second, when the unknown divergence time was inferred with a strict clock, lineage-specific rates were present, and 400 loci were sampled, only 11.5% of replicate experiments included the correct value within the 95% HPD (Table 2.3).

Given that the rate prior is rarely well informed, the sensitivity of the 95% HPD to the variance of the rate prior presents a serious problem for divergence time estimation. This problem is especially important when there are lineage-specific rates, because their impact is not reduced by sampling more loci. This contrasts to gene-specific rates and residual rates. For these classes of rate variation, sampling more loci reduces their impact. This in turn means the assumptions of the rate prior are less likely to be violated and the 95% HPD is more likely to include the correct value (Table 2.1, Table 2.3).

A further result from the experiments presented here is that if the variance of the rate prior is less than that of the distribution from which rate variation was simulated, sampling more data (either more loci or more bp at a single locus) can reduce the probability that the 95% HPD includes the correct value. This is most strikingly expressed when there are lineage-specific rates and the unknown divergence time is inferred with a strict clock. In this case, the percentage of experiments that include the correct value within the 95% HPD falls from 91% when 800 bp or 1 locus is sampled, to 13% or 11.5% when 320,000 bp or 400 loci are sampled (Table 2.3). Patterns such as this are most clearly observed when there are lineage-specific rates (Table 2.1 and Table 2.3). This is likely explained by the fact that sampling more data does not reduce the impact of this class of rate variation.

The implications of lineage-specific rates for divergence time estimation with empirical datasets

Sister taxa comparisons in the *Ipomoea* phylogeny indicated lineage-specific rates that are comparable to the parameters explored in the simulation experiments presented here. Further, it is likely that sister taxa comparisons within a plant genus underrepresent the extent of lineage-specific rates that occur at broader phylogenetic scales. Given that stronger lineage-specific rates lead to larger errors in divergence time estimates (Fig. 2.8, Table 2.4) lineage-specific rates may be even more problematic for divergence time estimation in deeper phylogenies.

It is difficult to determine precisely the implications of lineage-specific rates when inferring divergence times with empirical datasets. Most time-calibrated phylogenies are far more complex than the three or eight-taxon trees used in the simulations presented here. As well as containing more taxa, they typically incorporate temporal assumptions through the implementation of multiple fossil calibrations (Near and Sanderson 2004; Near et al. 2005; Yang and Rannala 2006; Warnock et al. 2011, 2015; Magallón et al. 2015) and a constant rate birth-death branching process (Yang and Rannala 1997). These assumptions interact directly with inferences of substitution rates, which in turn will have complex effects on the distribution of age estimates (Welch and Bromham 2005; Donoghue and Benton 2007; Magallón et al. 2013; Donoghue and Yang 2016). Further, divergence time estimates in empirical datasets may also be affected by model misspecification that is considerably

more complex than in the simulation experiments presented here (Gillespie 1991; Sanderson et al. 2004; Duchêne et al. 2014; Kspeka and Phillips 2015; Field et al. 2019).

None of these complexities are likely to ameliorate the basic finding of this study, that in the presence of lineage-specific rates, increasing either the number of sampled loci or sequence length at a single locus has a limited effect in reducing error in divergence time estimates. This finding is concerning when considered in the context of the other assumptions and sources of evidence used in divergence time estimation. For example: variances of UCLN relaxed clocks are often arbitrarily specified, the constant rate birth death branching process is often likely to be violated, and the fossil record can often be highly incomplete and provide a misleading temporal framework from which to derive fossil calibrations (Smith 2001; Bromham 2006; Donoghue and Yang 2016). In the future, models that explicitly account for the relationship between certain traits and lineage-specific rates may be used more widely to infer divergence times (Lartillot and Poujol 2011; Ho 2014; Berv and Field 2018). However, it is currently the case that these models are rarely used, and even if they become increasingly easy to implement, they will inevitably remain sensitive to differing interpretations of how traits evolve. Taken together, divergence time estimation therefore remains one of the most challenging inference problems in molecular phylogenetics, regardless of the quantity of molecular sequence data available.

3. ESTIMATING DIVERGENCE TIMES WITH MULTIPLE FOSSIL CALIBRATIONS AND A RELAXED CLOCK

ABSTRACT

Relaxed clock methods account for among-branch-substitution-rate-variation when estimating divergence times by inferring different rates for individual branches. In order to infer different rates for individual branches, important assumptions are required. This is because molecular sequence data do not provide direct information about rates, but instead provide direct information about the total number of substitutions along any branch, which is a product of the rate and time for that branch. Often, the assumptions required for estimating rates for individual branches depend heavily on the implementation of multiple fossil calibrations in a single phylogeny. Here, I show that the basis of these assumptions is often critically undermined. First, I highlight that the temporal distribution of the fossil record often violates key assumptions of methods that use multiple fossil calibrations with relaxed clocks. With respect to "node calibration" methods, this conclusion is based on the inference that different fossil calibrations are unlikely to reflect the relative ages of different clades. With respect to the fossilised-birth-death-process, this conclusion is based on the inference that the fossil recovery rate is often highly heterogeneous. I then demonstrate that methods of divergence time estimation that use multiple fossil calibrations are highly sensitive to assumptions about the fossil record and among-branch-substitution-rate-variation. Given the problems associated with these assumptions, the results of this study highlight that using multiple fossil calibrations with relaxed clocks often does little to improve the accuracy of divergence time estimates.

INTRODUCTION

In phylogenetics, molecular sequence data provides direct information about the number of substitutions that have occurred since homologous sequences diverged. The number of substitutions is a product of the rate that the sequences are evolving (substitution rate), and the time since they diverged (Zuckerkandl and Pauling 1962, 1965; Margoliash 1963). Additional evidence is therefore required to estimate substitution rates and times in molecular phylogenies. Often, this additional evidence is a fossil calibration, where the age of a fossil is used as a basis for estimating the age of a particular clade (Zuckerkandl and Pauling 1962; Benton 1995; Sanderson 1997; Donoghue et al. 2001; Renner 2004; Donoghue and Benton 2007; Gandolfo et al. 2008; Donoghue and Yang 2016). This provides a timeframe over which molecular evolution within the clade has occurred, enabling the estimation of a substitution rate. In the simplest case of the strict molecular clock, the rate is the same for every branch in a phylogeny (Zuckerkandl and Pauling 1962, 1965; Margoliash 1963; Miyata 1980; Baldwin and Sanderson 1998). A single well-placed calibration can therefore enable the inference of accurate divergence times throughout the phylogeny.

Often, rates vary among branches (Fig. 3.1) (Langley and Fitch 1974; Britten 1984; Gillespie 1989, 1991; Bromham et al. 1996). Relaxed clock methods that allow among-branch-substitution-rate-variation are therefore often used (Sanderson 1997, 2002; Thorne et al. 1998; Kishino et al. 2001; Drummond et al. 2006; Drummond and Suchard 2011; Lartillot et al. 2016). These methods rely heavily on assumptions about rates and times of individual branches. Often, multiple fossil calibrations are used in a single phylogeny as a basis for making these important assumptions, providing "landmarks" to calibrate the relaxed clock (Sanderson 1997; Donoghue and Benton 2007; Magallón et al. 2015; Donoghue and Yang 2016). The use of multiple fossil calibrations to provide "landmarks" to calibrate the relaxed clock – and facilitate the estimation of rates and times for individual branches, and therefore the relative ages of different clades within a single phylogeny – represents a fundamental change from a traditional view of the fossil record, where fossil ages are interpreted purely as minimum age estimates for their respective clades (Donoghue and Benton 2007).

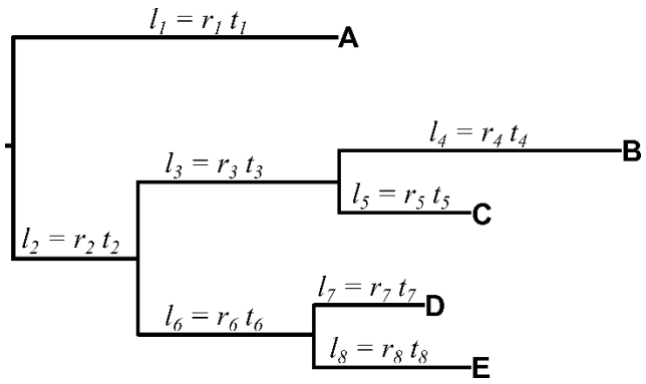

Figure 3.1. Branch lengths (l_{1-8}) are equal to the number of substitutions that have occurred on each branch. l_{1-8} are a product of the substitution rate for each branch (r_{1-8}) and the temporal duration of each branch (t_{1-8}). Only l_{1-8} can be directly inferred from molecular sequence data. In the case of the strict molecular clock, r_{1-8} are equal. When rates vary among branches, r_{1-8} are not equal.

Previous studies have evaluated the robustness of methods that use multiple fossil calibrations and relaxed clocks (Sanderson 1997, 2002; Thorne et al. 1998; Kishino et al. 2001; Near and Sanderson 2004; Britton 2005; Near et al. 2005; Drummond et al. 2006; Yang and Rannala 2006; Marshall 2008; Drummond and Suchard 2011; Magallón et al. 2013; Gavryushkina et al. 2014, 2017; Heath et al. 2014; Warnock et al. 2015, 2017; Zhu et al. 2015; Lartillot et al. 2016; Zhang et al. 2016). These studies have provided critical insights into the performance of different methods, and the implications of different assumptions about the nature of the fossil record and molecular evolution. Such studies typically consider one of either the fossil record (Near and Sanderson 2004; Near et al. 2005; Yang and Rannala 2006; Marshall 2008; Warnock et al. 2011, 2015, 2017; Heath 2012; Gavryushkina et al. 2014, 2017; Heath et al. 2014; Zhang et al. 2016) or molecular evolution (Thorne et al. 1998; Kishino et al. 2001; Drummond et al. 2006; Drummond and Suchard 2011; Zhu et al. 2015; Lartillot et al. 2016), and rarely consider the interactions between these two types of evidence (Magallón et al. 2013). This is despite the fact that such interactions are likely to be important in divergence time estimation – assumptions about the fossil record affect assumptions about times, assumptions about molecular evolution affect assumptions about rates, and the product of the inferred times and rates is the number of substitutions, the parameter directly inferred from molecular sequence data.

Further, the majority of methodological studies of fossil calibrations are performed in groups with well-preserved fossil records, or where fossil calibrations can be implemented in numerous clades throughout the phylogeny (Near and Sanderson 2004; Near et al. 2005; Warnock et al. 2011, 2015, 2017; Heath 2012; Magallón et al. 2013; Gavryushkina et al. 2014, 2017; Heath et al. 2014; Zhang et al. 2016). In many empirical studies that include divergence time estimates, the fossil record of the relevant group is considerably more fragmentary and therefore provides an extremely incomplete picture of the temporal range and diversity of a given taxon (as is the case across the vast majority of taxa in the tree of life) (for example Särkinen et al. 2013; Lagomarsino et al. 2016; Mitchell et al. 2016; Cardillo et al. 2017; Folk et al. 2019).

Here, I explore the validity and implications of different assumptions about the fossil record and molecular evolution, and also the interaction between these two types of evidence, in Osmundaceae, and Convolvulaceae and Solanaceae (CS). Time-calibrated phylogenies have recently been constructed for these two clades according to commonly used methodologies (Parham et al. 2012; Särkinen et al. 2013; Grimm et al. 2015; Mitchell et al. 2016). The analyses presented here can therefore be interpreted in the context of commonly used methodologies. Further, these clades have markedly contrasting fossil records with Osmundaceae possessing a far better preserved fossil record than CS (Särkinen et al. 2013; Bomfleur et al. 2015; Grimm et al. 2015). Key findings can therefore be compared in different contexts, unlike the majority of previous studies that are carried out in groups with well-preserved fossil records (Near and Sanderson 2004; Near et al. 2005; Warnock et al. 2011, 2015, 2017; Heath 2012; Magallón et al. 2013; Gavryushkina et al. 2014, 2017; Heath et al. 2014; Zhang et al. 2016). As well as these two clades, I also perform a subset of analyses across the entire of Spermatophyta. A time-calibrated phylogeny has recently been constructed for this clade that incorporates over 100 fossil calibrations, and which has been used as a reference for inferring divergence times in hundreds of subsequent studies (Magallón et al. 2015). The results of analyses presented here can therefore be interpreted in the context of a study of much broader phylogenetic scale that incorporates a very large number of fossil calibrations.

In the analyses presented here, I first investigate whether the temporal distribution of the fossil records of these clades is consistent with the assumptions of methods that use multiple fossil calibrations with relaxed clocks. Specifically, I explore whether the relationship between the ages of sets of fossils used as calibrations throughout a phylogeny, and the actual ages of the respective clades of these fossils, provides a framework for estimating rates and times on individual branches, and thus the relative ages of different clades. I explore the nature of this relationship in the context of the assumptions of node calibration methods and the fossilised-birth-death-process (FBDP).

To investigate whether the temporal distribution of the fossil record is consistent with the assumptions of methods that use multiple node calibrations, I estimate whether the ages of sets of fossils previously used together as part of a set of fossil calibrations in a single phylogeny – subsequently referred to as *node-calibration-fossils* – reflect the relative ages of different clades (Fig. 3.2a), or instead have no such relationship (Fig. 3.2b). I focus on this, because to provide relevant information about times and rates for individual branches, sets of node-calibration-fossils must provide information about relative clade ages. In contrast to this, the highly variable time-lags between a set of node-calibration-fossils and the actual ages of their respective clades may be such that a given set of node-calibration-fossils may not accurately reflect the relative ages of clades throughout the phylogeny.

It is fundamentally important that the ages of sets of node-calibration-fossils reflect relative clade ages, even though strictly speaking, the ages of these fossils typically represent the minimum constraints of a fossil calibration. In a many studies such as those discussed here (Särkinen et al. 2013; Grimm et al. 2015; Magallón et al. 2015), the fossil age is the only physical evidence from which a calibration is derived, with other characteristics of the fossil calibration being derived arbitrarily. Because of this, it is often the case that the age of the fossil directly affects the temporal assumptions expressed through the entire fossil calibration (for example Särkinen et al. 2013; Grimm et al. 2015; Magallón et al. 2015). As such, in order for the fossil calibrations to reflect relative clade ages (and

enable the implementation of a relaxed clock) sets of node-calibration-fossils must provide some information with respect to relative clade ages.

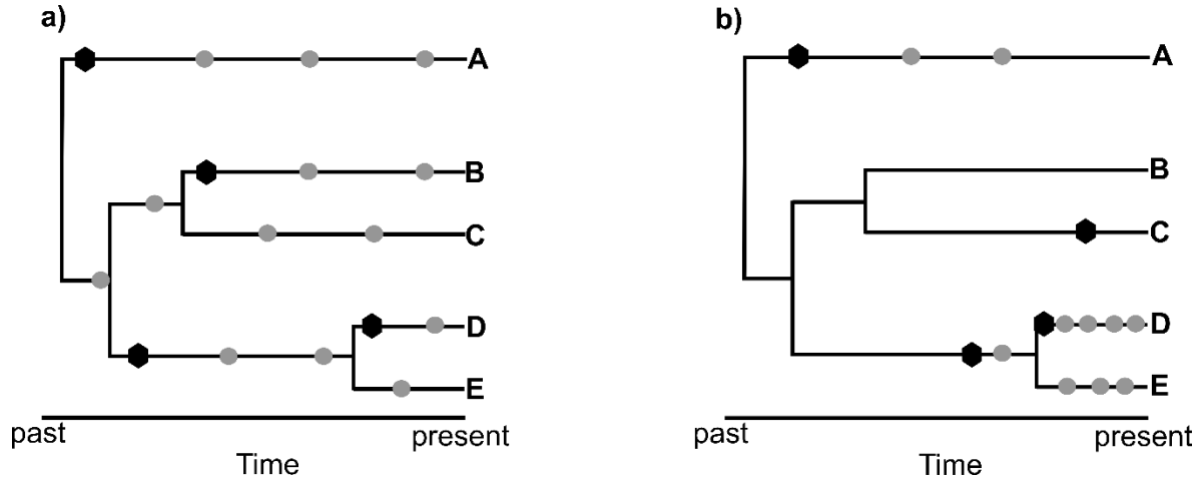

Figure 3.2. A summary of two contrasting fossil records. Fossils are sampled along the branches of a macroevolutionary process that represents the "true" evolutionary history of the extant species A-E. The oldest fossil in each clade is a black hexagon. All other fossils are grey circles. In a) the oldest fossils in each clade (node-calibration-fossils) reflect the relative ages of different clades, and the fossil sampling rate (ψ) is constant. In b) the oldest fossils in each clade do not reflect the relative ages of different clades, and ψ is heterogeneous.

I investigate whether sets of node-calibration-fossils meet this criteria in two different ways. First, I analyse the temporal distribution of all known fossils within their respective clades, and use methods derived from palaeontology for estimating taxon age ranges (Marshall 1990, 2008; Springer 1995) to calculate confidence intervals for the ages of these clades. This provides a basis to quantify the relationship between the ages of node-calibration-fossils and the actual ages of their respective clades. Second, I calculate the empirical scaling factor (ESF) for node-calibration-fossils following the method of Marshall (2008). The ESF provides an alternative approximation of how the age of a node-calibration-fossil reflects the age of its respective clade by taking into account the age of a fossil and the phylogenetic depth of the node that it calibrates (see methods). By quantifying how the ages of node-calibration-fossils reflect actual clade ages, both of these measures (confidence intervals and ESFs) provide a basis to determine how sets of node-calibration-fossils are likely to reflect the relative ages of different clades.

To investigate whether the temporal distribution of the fossil record is consistent with the assumptions of methods that use the FBDP, I analyse the fossil-recovery-rate (ψ). This parameter describes the rate that fossils are sampled from a branching process. I specifically focus on determining whether ψ is constant (Fig. 3.2a), or instead whether it is heterogeneous (Fig. 3.2b). This is because a constant ψ is *the* fundamental assumption that the FBDP makes of the temporal distribution of the fossil record, and in groups for which morphological matrices are not available for relevant fossils – as is often the case in plants – ψ is the single parameter through which fossil ages are incorporated into divergence time analyses (Stadler 2010; Gavyryushkina et al. 2014; Heath et al. 2014).

Following this investigation of the temporal distribution of the fossil record, I then infer divergence times for Osmundaceae and CS in different methodological contexts. The sensitivity of different methods to the assumptions they make about the temporal distribution of the fossil record can therefore be determined. For this second set of analyses, I compare methods for implementing fossil calibrations – subsequently referred to as *calibration strategies*, and methods for analysing molecular sequence data – subsequently referred to as *molecular clock strategies.* I compare four different node calibration strategies. *Non-Uniform Precise* (NUP) and *Non-Uniform Imprecise* (NUI) use non-uniform probability densities (calibration densities) that assume the clade age is very similar to the age of the node-calibration-fossil (NUP) (Fig. 3.3a), or make less precise assumptions about the clade age in relation to the node-calibration-fossil (NUI) (Fig. 3.3b). *Uniform Precise* (UP) and *Uniform Imprecise* (UI) use uniform calibration densities in which the maximum is either young (UP) (Fig. 3.3c) or old (UI) (Fig. 3d). For calibration strategies, I also implement the FBDP. For molecular clock strategies, I compare; a *strict clock* – where the same rate is inferred for every branch; a *relaxed clock* – where rates are inferred independently for each branch; and no *molecular data* – where no molecular data is analysed when inferring divergence times.

It is important to note that the purpose of this second set of analyses is not to estimate the most accurate timescale for the evolution of these two groups. Instead, the purpose is to provide a comparative framework to compare the implications of different methodological assumptions. In

order to provide such a comparative framework, I inevitably implement some methodologies, such as the strict clock, that are no longer widely used. As well as implementing methodologies that are no longer widely used, I also use terminology that best reflects the comparative framework that is central to this study. For example, I refer to no molecular data as a molecular clock strategy, rather than the effective time prior. This is because the purpose of different molecular clock strategies is to compare the implications of different methods for analysing molecular sequence data. I consider no molecular data as an extreme case, and a baseline from which to evaluate alternative approaches for analysing molecular sequence data.

When comparing inferred divergence times with these different methods, I refer to mean posterior age estimates (MPEs) and 95% highest posterior density (HPD) intervals. I also compare the difference between the age of the oldest fossil for a clade (the node-calibration-fossil for node calibration methods), and the MPE for that clade. This quantifies the relationship between fossil ages and inferred divergence times. The comparisons between these different methods enable a comprehensive analysis of the interaction between different assumptions about the fossil record and molecular evolution, and the implications of this for divergence time estimation.

Taken together, the analyses carried out in this chapter are designed to set out the extent to which the fossil record is consistent with the assumptions of methods that use multiple fossil calibrations with relaxed clocks. This context is then used as a basis for a comprehensive analysis of the implications of different assumptions about molecular evolution and the fossil record when estimating divergence times, and for evaluating the extent to which inferences are sensitive to underlying assumptions.

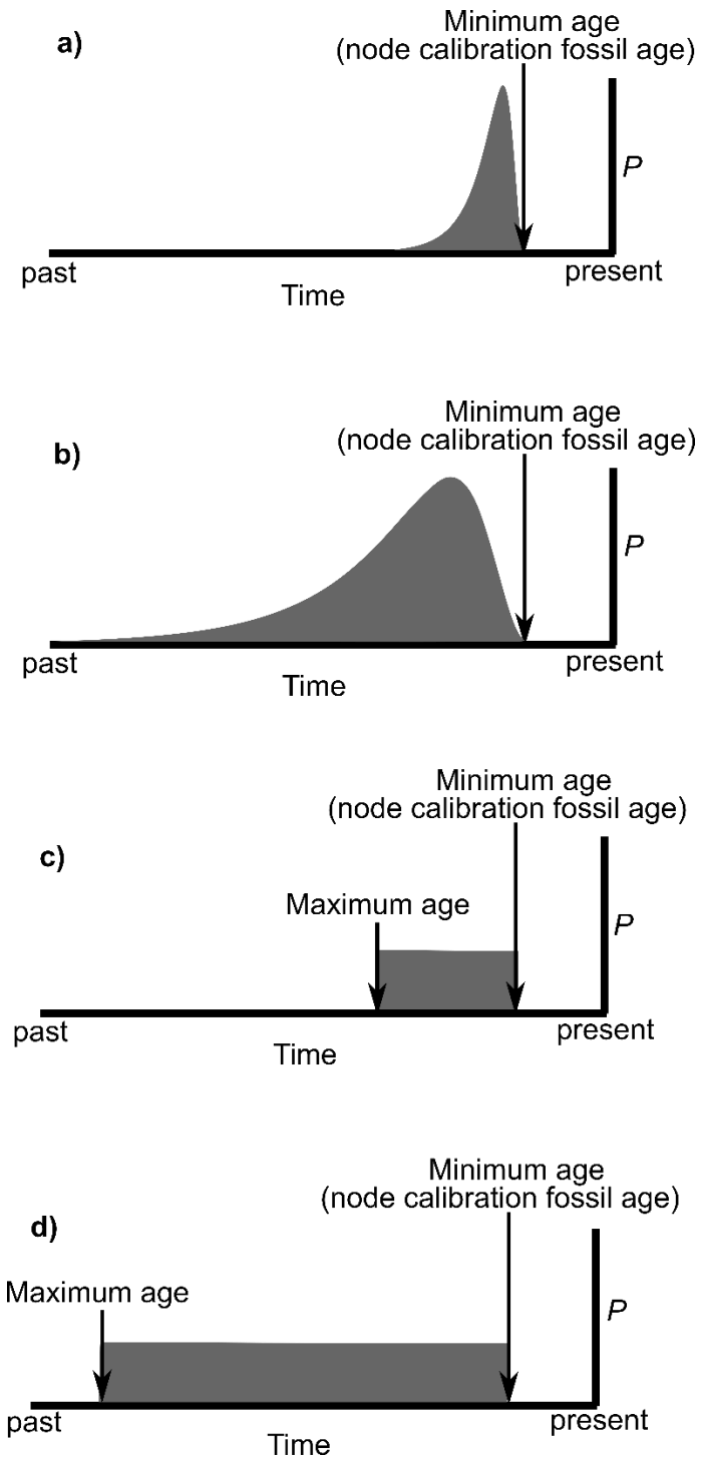

Figure 3.3. A summary of the different node calibration strategies that are compared in this study. a) Non-Uniform Precise – a non-uniform calibration density assumes the actual age of the clade is similar to the age of the node-calibration-fossil. b) Non-Uniform Imprecise – a non-uniform calibration density makes less precise assumptions about the age of the clade in relation to the age of the node-calibration-fossil. c) Uniform Precise – a uniform calibration density that assumes the age of the clade is similar to the age of the node-calibration-fossil. d) Uniform Imprecise – a uniform calibration density that makes less precise assumptions about the age of the clade in relation to the age of the node-calibration-fossil.

MATERIALS AND METHODS

Molecular Data

The molecular matrix for Osmundaceae is the same as that used by Grimm et al. (2015) and originally compiled by Metzgar et al. (2008). It consists of 13 taxa sampled for three chloroplast genes (*rbcL*, *atpA* and *rps4*) and five spacer regions (*rps4-trnS, trnG-trnR, trnL-trnF, rbcL-accD, atpB-rbcL*). The matrix for CS contains 56 taxa sampled for four chloroplast genes (*atpB*, *matK*, *rbcL* and *ndhF*), the chloroplast intergenic spacer *trnL-trnF,* and the nuclear marker *ITS.* Apart from species of *Ipomoea*, for which sequences from Muñoz-Rodríguez et al. (2019) were used, these sequences were downloaded from GenBank (Appendix 1). The matrix for Spermatophyta is based on that of Magallón et al. (2015). It consists of 303 taxa sampled for two nuclear genes (*18S* and *26S*) and three chloroplast genes (*matK, rbcL* and *atpB*) (Appendix 1).

Fossil Data

For Osmundaceae, a dataset of 19 rhizome fossils and 17 frond fossils was analysed (Appendix 2). Grimm et al. (2015) originally compiled this dataset, and assigned these fossils to one of five clades: Osmundaceae, *Osmunda*, *Osmunda japonica – Osmunda vachellii* (JV), *Osmunda japonica – Osmunda regalis* (JR), and *Todea barbara – Leptopteris wilkesiana* (BW). The 19 rhizome fossils possess sufficient morphological characters to be included in a 33-character morphological matrix (Bomfleur et al. 2015).

For CS, a dataset of 32 seed fossils originally compiled by Särkinen et al. (2013), was analysed (Appendix 2). Särkinen et al. (2013) assigned these fossils to either Solanaceae (including the stem branch), or Solanoideae (including the stem branch). Four additional fossils have putatively been assigned to Convolvulaceae (MacGintie 1954; Martin 2000; Martin 2001; Mitchell et al. 2016; Srivastava et al. 2018), and a further fossil that is drastically older than previously known fossils has recently been assigned to Solanoideae (Wilf et al. 2017). These fossils were not included in this study for two reasons. First, they cannot be conclusively assigned to specific clades within CS. Second, the purpose of this study is to investigate the implications of fossil calibrations in a way that is directly

relevant and comparable to methods implemented in previous studies. These fossils have not been used as calibrations in previous studies.

For Spermatophyta, a dataset of 113 fossils that is based on that of Magallón et al. (2015), was analysed. This dataset contains fossils that belong to clades throughout Spermatophyta (Appendix 2).

Phylogenetic inference

Molecular phylogenies were constructed for sampled extant taxa. These would serve as starting topologies for subsequent analyses. For Osmundaceae, the same alignment was used as Grimm et al. (2015). For CS and Spermatophyta, sequences were aligned with MAFFT v7.271 (Katoh 2002; Katoh and Standley 2013), using the L-INS-I setting and a gap opening penalty of 1.53. Ambiguously aligned regions were removed using default settings in Gblocks (Castresana 2000; Talavera and Castresana 2007) and aligned sequences for each marker were concatenated using Sequence Matrix v1.8 (Vaidya et al. 2011).

mrBayes v3.2.6 (Huelsenbeck and Ronquist, 2001; Ronquist et al. 2012) was used for phylogenetic analyses. A GTR + G + I model of sequence evolution was used. In all analyses two independent runs were performed. For Osmundaceae and CS, the analysis was run for 2,000,000 generations (sampling every 200 generations). The final split frequency between runs was < 0.01 in both cases. For Spermatophyta, the analysis was run for 4,500,000 generations (sampling every 2,000 generations). The final split frequency between runs was 0.0185.

Confidence intervals for clade ages

For Osmundaceae and CS, the fossil datasets represent the entire known fossil record (Särkinen et al. 2013; Grimm et al. 2015). Confidence intervals for the ages of clades within these two groups could therefore be calculated using approaches analogous to previously developed methods for estimating taxon ages (Marshall 1990, 2008; Springer 1995). For Osmundaceae, confidence intervals were calculated for: the entire of Osmundaceae, *Osmunda*, and JV. Confidence intervals could not be calculated for JR or BW because only one fossil was sampled from these clades. For CS, confidence

intervals were calculated for the entire of Solanaceae (including the stem branch) and Solanoideae (including the stem branch).

To calculate confidence intervals, Kolmogrov-Smirnov tests were first performed to determine the distribution to which fossil occurrence times within these two clades conform. Two distributions were tested: a uniform distribution in which the minimum was the present and the maximum was the age of the oldest fossil, and an exponential distribution for which the mean was equal to the mean age of all fossils within the clade.

The best fitting distribution is interpreted as a description of how fossils are sampled through time from a given clade. An interval that is older than the oldest fossil for a clade, and within which there would be a 95% probability (according to the inferred distribution) of sampling a fossil were the clade still in existence, represents a 95% confidence interval for the actual clade age.

For the uniform distribution, the upper limit of the 95% confidence interval for the actual clade age (CI_{max}) is given by: $CI_{max} = \frac{OF}{\sqrt[n]{1-p}}$. OF is the age of the oldest fossil, n is the number of fossils and p is the confidence level (in this case 0.95). For the exponential distribution, CI_{max} is given by: $CI_{max} = OF + Q(p)$. $Q(p)$ is the quantile function of the part of the exponential distribution that is older than OF. p is the confidence level (in this case 0.95).

When calculating confidence intervals, uncertainty in fossil ages was accounted for. Using a custom R script, 10,000 replicate calculations were performed. In each replicate, the age of each fossil was a drawn randomly from its stratigraphic range. When plotting the upper bound of the 95% confidence interval, the thickness of the bar represents the interval between the 2.5th and 97.5th percentiles following 10,000 replicate calculations.

When analysing the fossil record of CS, the high proportion of Pliocene fossils made fitting a distribution problematic. Additional analyses were therefore performed where a subset of the Pliocene fossils was sampled. Appropriate distributions were inferred and confidence intervals were calculated based on this subsample. This subsampling approach is likely to be robust for two reasons. First, these young Pliocene fossils are unlikely to provide useful information about the ages of clades within CS

because they are far younger than many other fossils within their respective clades. Second, by removing these fossils, we could more accurately model the temporal distribution of older fossils within their respective clades. Regardless, to ensure that inferences were robust, the entire fossil dataset was also used to infer distributions and calculate confidence intervals.

Calculating ESFs for node-calibration-fossils

For all three datasets, ESFs were calculated for node-calibration-fossils. This followed the approach of Marshall (2008). With FA being the absolute age of the fossil, and D being the depth of the node in the ultrametric phylogeny to which the node-calibration-fossil is assigned, $ESF = \frac{FA}{D}$.

The ultrametric phylogenies necessary for calculating ESFs were inferred in the penalised likelihood framework treePL (Smith and O'Meara 2012). The trees inferred in mrBayes v3.2.6 were used as input trees. Cross-validation was used to determine the optimum smoothing value. The smoothing value describes the extent to which rate differences between ancestral and descendant branches are penalised when inferring the ultrametric phylogeny, with a higher smoothing value penalising rate differences more than a lower value. For Osmundaceae a smoothing value of 10,000 was used, for CS a smoothing value of 0.01 was used, and for Spermatophyta a smoothing value of 0.01 was used.

The rationale for comparing ESFs rests on the assumption that the relative *depths* of different nodes in an ultrametric phylogeny reflect the relative *ages* of different clades. If this is the case, a comparison of the age of a set of node-calibration-fossils relative to the depth of the node to which they are assigned ($\frac{FA}{D}$) provides a measure of variation in how different node-calibration-fossils reflect the actual age of their respective clade.

In order for the relative depths of different nodes to reflect the relative ages of different clades, rate variation must be estimated accurately when inferring the ultrametric phylogeny. Given molecular sequence data only provides direct information about the expected number of substitutions on each branch (expressed in the branch length estimates for the input trees in treePL), this is a

problematic requirement, and methods for constructing ultrametric phylogenies are likely to be sensitive to underlying assumptions.

In this study, ultrametric phylogenies were inferred in treePL because it does not implement the constant rate birth-death branching process (characteristic of Bayesian analyses), that may, through the assumptions it makes about times, significantly bias inferences of rate variation. This problem may be especially important in the large Spermatophyte dataset. Regardless, the assumption of rate variation made by treePL, whereby rates are inherited between ancestral and descendant branches, is one of potentially many assumptions that could be made about rate variation. Analyses of ESFs must therefore be treated with caution, but in the context of other analyses, such as the calculation of confidence intervals for clade ages, can provide important insights.

Calculating the fossil recovery rate ψ

The ultrametric phylogenies were also used as a basis to estimate ψ within Osmundaceae and CS. Depending on the phylogenetic placement of fossils, ψ was inferred within different clades within these two groups, but in some cases also within different grades. With n being the number of fossils assigned to a clade or grade, and BL being the total length of all branches within the clade or grade in the ultrametric phylogeny, $\psi = \frac{n}{BL}$.

This calculation of ψ is directly relevant to how ψ is formalised in the FBDP – where it describes the rate that fossils are sampled along the branches of the macroevolutionary process that generated sampled extant taxa and sampled fossils. In some formalisations of the FBDP, fossils are assigned to more specific parts of the tree than they are here. Further, the effect of sampling fossils along extinct branches is not explicitly considered here. However, the estimates derived from these calculations provide a useful approximation of the extent to which ψ is likely to vary.

Inferring divergence times with different molecular clock strategies and different calibration strategies

Node calibrations

For Osmundaceae, four methods for implementing node calibrations were used: NUP, NUI, UP and UI. NUP was the same as the strategy used by Grimm et al. 2015. Each of the five fossil calibrations was implemented as an exponential distribution. The offset was equal to the minimum age of the stratum containing the node-calibration-fossil, and the 97.5th percentile was equal to the maximum age of the stratum. In NUI, each of the five fossil calibrations was implemented as a lognormal distribution. The offset was equal to the minimum age of the stratum containing the node-calibration-fossil, the mean (μ) was parameterised such that the 95th percentile of the calibration density equals the upper limit of the 95% confidence interval that was calculated for the clade's age, and the standard deviation (σ) was 1. In UP, the fossil calibrations were implemented as uniform distributions. The minimum was equal to the minimum age of the stratum containing the node-calibration-fossil, and the maximum for all calibrations was 299 Ma. This corresponds to a previously estimated divergence time between osmundaceous ferns and other leptosporangiate ferns (Schuettpelz and Pryer 2009). In UI, the fossil calibrations were also implemented as uniform distributions. The minimum age was equal to the minimum age of the stratum containing the node-calibration-fossil, and the maximum for all calibrations was 472 Ma. This corresponds to the age of the oldest known cryptospores (Rubinstein et al 2010).

For CS, four methods for implementing node calibrations were also used: NUP, NUI, UP, and UI. NUP was the same as the strategy used by Särkinen et al. (2013). The two fossil calibrations were implemented as lognormal distributions. The offset was equal to the minimum age of the stratum containing the node-calibration-fossil, $\mu = 0.01$, and $\sigma = 1$. NUI was parameterised according to the same approach as Osmundaceae, but the confidence intervals for clade ages were used that were calculated when 13 pliocene fossils were removed. In UP, fossil calibrations were implemented as uniform distributions. The minimum was equal to the minimum age of the stratum containing the node-calibration-fossil, and the maximum for all calibrations was 86.9 Ma. This corresponds to the

upper limit of the 95% HPD for the divergence time between Convolvulaceae and Solanaceae in Magallón et al. (2015). In UI, fossil calibrations were also implemented as uniform distributions. The minimum was equal to the minimum age of the stratum containing the node-calibration-fossil, and the maximum for all calibrations was 130 Ma. This corresponds to the maximum age of the stratum in which tricolpate pollen appears in the fossil record (Doyle et al. 1977). I used this is a maximum constraint for UI, because aside from secondary calibrations derived from previous studies (eg Magallón et al. 2015), there is no robust maximum that can be applied to the divergence time between these two clades.

Time-calibrated phylogenies were inferred in RevBayes. The topology was fixed to that previously inferred in mrBayes v3.2.6. A GTR + G + I model of sequence evolution was used in all cases. Analyses were performed with a strict clock, UCLN relaxed clock, and no molecular data. A constant rate birth-death branching process was used as the tree prior. For each analysis, two independent runs were performed. Sufficient mixing and convergence was assessed in Tracer v1.6.0 (Rambaut et al. 2014). A 25% burn-in was used prior to calculating mean posterior estimates (MPEs), and 95% highest posterior density intervals (HPDs).

FBDP

FBDP analyses were performed for Osmundaceae. These analyses included all 32 Osmundaceae fossils, and the 33-character morphological matrix for the 19 rhizome fossils. The frond fossils were incorporated by using the unresolved FBDP (Heath et al. 2014). This approach takes into account uncertainty in the phylogenetic position of the frond fossils, whilst allowing the analysis to be constrained to reflect existing knowledge of their phylogenetic position. The FBDP analyses were performed in RevBayes. A GTR + G + I model of sequence evolution was used for the molecular data, with either a strict molecular clock or a UCLN relaxed clock. A symmetrical Markov model was used for the morphological data, with a strict clock that assumes morphological characters evolve at the same rate on all branches. For each analysis, two independent runs were performed. Sufficient mixing and convergence was assessed in Tracer v1.6.0. A 25% burn-in was used prior to calculating mean posterior estimates (MPEs), and 95% highest posterior density intervals (HPDs).

An analysis with no molecular data was not performed with the FBDP because it does not represent an appropriate or useful critique of the FBDP. The FBDP is highly integrated, with the times and morphological characters of fossils, and molecular sequence data of extant taxa, being modelled in a single macroevolutionary framework. As such, it would be difficult to analyse the precise effects of not including molecular sequence data.

Implementing the FBDP in CS was considered. This would necessarily use the unresolved FBDP because there is no appropriate morphological matrix available for known fossils (Heath et al. 2014). However, in initial experiments, I found that implementing this method with this dataset was problematic because independent runs did not converge – even following extensive experimentation with different mixing procedures for the mcmc algorithm. Future developments of this new method may enable it to function more effectively when analysing large numbers of fossils with limited morphological characters.

Custom R and RevBayes scripts that were developed for analyses in this study are available in Electronic Appendix 2.

RESULTS

Phylogenetic distribution of fossils

Phylogenies for Osmundaceae, CS, and Spermatophyta were inferred as a basis for subsequent analyses. These phylogenies are congruent with previous studies (Appendix 3) (Särkinen et al. 2013; Grimm et al. 2015; Magallón et al. 2015). Fossils in this study are therefore assigned to the same clades as previous studies (Appendix 2) (Särkinen et al. 2013; Grimm et al. 2015; Magallón et al. 2015).

The fossils analysed in Osmundaceae and CS are the entire known fossil records for these groups (Särkinen et al. 2013; Grimm et al. 2015). Figure 3.4a summarises the phylogenetic distribution of fossils within Osmundaceae. 36 fossils belong to Osmundaceae, 27 fossils belong to *Osmunda*, 16 fossils belong to the clade of *O. japonica* – *O. vachellii* (JV), 1 fossil belongs to the

clade of *O. japonica – O. regalis* (JR), and 1 fossil belongs to the clade of *T. barbara – L. wilkesiana* (BW). Figure 3.4b summarises the phylogenetic distribution of fossils within CS. 32 fossils belong to Solanaceae (including the stem lineage), 25 fossils belong to Solanoideae (including the stem lineage), and no fossils belong to Convolvulaceae (including the stem lineage).

For Spermatophyta, only the oldest known fossil within each clade was analysed. These fossils are distributed throughout Spermatophyta (Appendix 2).

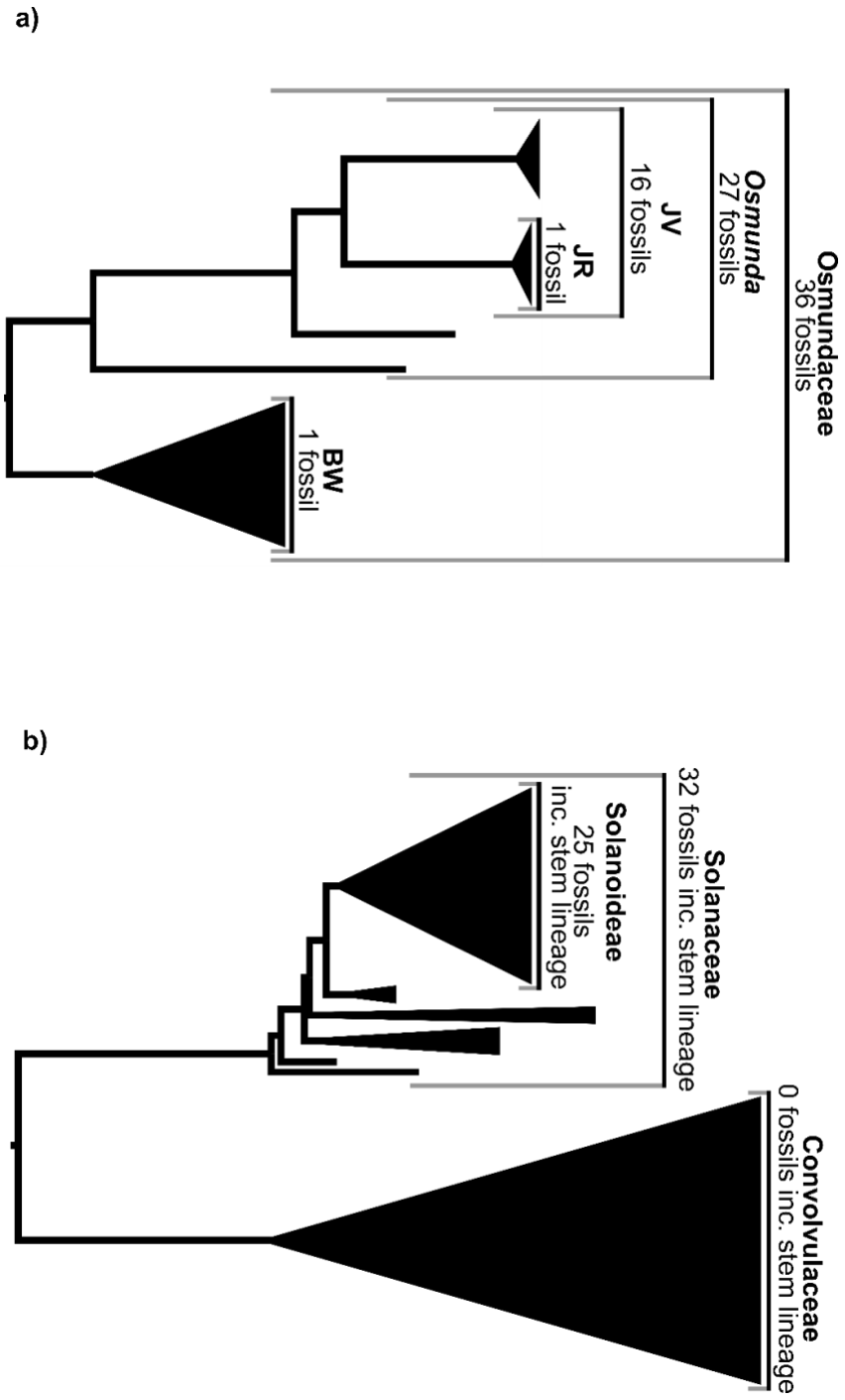

Figure 3.4. Summary of inferred phylogenies and number of sampled fossils within each clade, for the two datasets in which sampled fossils represent the entire known fossil record. a) Osmundaceae, b) CS.

Confidence intervals for clade ages

Osmundaceae

The fossil record for Osmundaceae extends into the Middle Triassic (237 Ma), for *Osmunda* it extends into the Middle Jurassic (165 Ma), for JV it extends into the Upper Cretaceous (86 Ma), for JR it extends into the Miocene (14 Ma), and for BW it extends into the Eocene (52 Ma) (Fig. 3.5a). For all clades with more than one fossil (Osmundaceae, *Osmunda* and JV), Kolmogrov-Smirnov tests indicate that fossils are sampled at an exponentially decreasing frequency with time from present (Table 3.1). Based on these distributions, the upper limit of a 95% confidence interval that includes the actual age of Osmundaceae is in the Lower Orodovician (471 Ma), for *Osmunda* it is in the middle Carboniferous (318 Ma), and for JV it is in the Lower Jurassic (197 Ma) (Fig. 3.5a, Table 3.2). These ages are far older than the 95th percentiles of calibration densities used in a recent time-calibrated phylogeny (Grimm et al. 2015) (Fig. 3.5a).

CS

The Solanaceae fossil record (including the stem lineage) extends into Eocene (48.0 Ma), and the Solanoideae fossil record (including the stem lineage) extends into the Oligocene (28.4 Ma) (Fig. 3.5b). For both clades, Kolmogrov-Smirnov tests indicate that fossils are more likely to be sampled at an exponentially decreasing frequency with time from present, rather than a uniform frequency (Table 3.3). However, the high frequency of Pliocene fossils made fitting an exponential distribution problematic. A subsample of the fossil record in which 13 Pliocene fossils were removed was therefore analysed. In this case, the exponential distribution was a far better fit for both clades (Table 3.3). When using this subsample of fossils, the upper limit of a 95% confidence interval that includes the actual age of Solanaceae (and its stem lineage) is in the Lower Cretaceous (104.9 Ma), and for Solanoideae (and its stem lineage) is in the Eocene (52.5 Ma) (Fig. 3.5b, Table 3.4). When Pliocene fossils were not removed, the upper limits of 95% confidence intervals were in the Upper Cretaceous (87 Ma) for Solanaceae (and its stem lineage) and still in the Eocene (45 Ma) for Solanoideae (and its

stem lineage) (Fig. 3.5b). Both sets of ages are far older than the 95th percentile of calibration densities used in a recent time-calibrated phylogeny (Särkinen et al. 2013) (Fig. 3.5b).

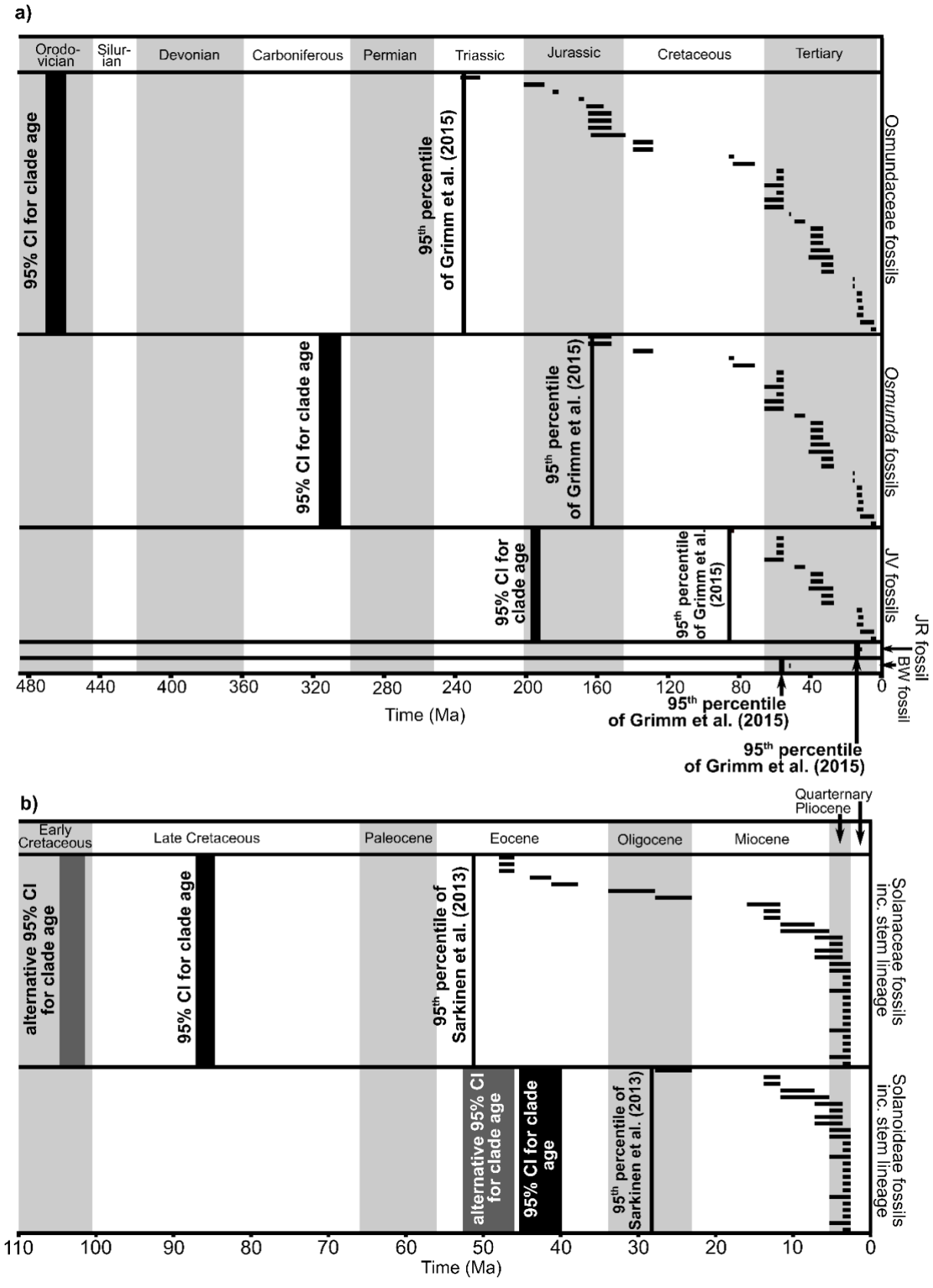

Figure 3.5. *(previous page)* The temporal distribution of sampled fossils in a) Osmundaceae, and b) CS. Black plotted points refer to individual fossil occurrences. Fossils are grouped into subclades – the same subclades as summarised in Figure 3.4. Where a clade has more than one sampled fossil, the upper bound of a 95% confidence interval (CI) for the actual clade age is indicated. This is labelled as the "95% CI for clade age". The width of this bar reflects uncertainty in fossil ages. For CS, two confidence intervals are calculated for each clade. Either the entire dataset of fossils is used when calculating the confidence interval, or a subsample is used in which some Pliocene fossils are removed. Where fossils have been removed, the relevant confidence interval is labelled as the "alternative 95% CI for clade age". The 95th percentile of calibration densities implemented in previous studies (Särkinen et al. 2013, Grimm et al. 2015) is also indicated. These are labelled as either "95th percentile of Grimm et al. (2015)" or "95th percentile of Särkinen et al. (2013)".

Table 3.1. Kolmogrov-Smirnov tests for the fossil record in Osmundaceae

Clade	P values (does the distribution of fossil occurrences differ significantly from specified distribution)	
	Exponential	Uniform
Osmundaceae	0.6	0.00008
Osmunda	0.5	0.00006
JV	0.3	0.2

Table 3.2. Summary information for different node-calibration-fossils within Osmundaceae

Node	Fossil age	Upper limit of 95% confidence interval for actual clade age (percentage difference relative to fossil age)	Fossil ESF
Osmundaceae	227	471 (+51.8)	227.0
Osmunda	153	318 (+107.8)	180.7
JV	84	197 (+155.8)	214.9
BW	52	--	90.0
JR	12	--	485.6

Table 3.3. Kolmogrov-Smirnov tests for the fossil record in CS

Clade	P values (does the distribution of fossil occurrences differ significantly from specified distribution)		
	Exponential	Uniform	Exponential (subsample)
Solanaceae (including stem lineage)	0.02	0.00000002	0.7
Solanoideae (including stem lineage)	0.001	0.00000004	0.2

Table 3.4. Summary information for different node-calibration-fossils within CS

Node	Fossil age	Upper limit of 95% confidence interval for actual clade age (percentage difference relative to fossil age)	Fossil ESF
Solanaceae (stem)	46	104.9 (+56.1)	46.0
Solanoideae (stem)	23	52.5 (+56.2)	99.9

ESFs of node-calibration-fossils

Osmundaceae

ESFs range from 90.0 to 485.6 (Fig. 3.6a, Table 3.2). This indicates a more than five-fold difference in how the ages of node-calibration-fossils reflect the actual age of their respective node. For example, assuming the age of the JR node-calibration-fossil (ESF = 485.6) is equal to the actual age of its node (solely for the purposes of explanation), the age of the BW node-calibration-fossil (ESF = 90.0) is 18.5% of the age of its node.

CS

The ESF of the node-calibration-fossil at the Solanaceae stem node is 46.0, and the ESF of the node-calibration-fossil at the Solanoideae stem node is 99.9 (Fig. 3.6b, Table 3.4). This indicates a more than two-fold difference in how the ages of these two node-calibration-fossils reflect the actual age of their respective node.

Spermatophyta

For Spermatophyta, the largest ESF is 790.1, and the smallest ESF is 26.9 (Fig. 3.6c). This indicates a more than 29-fold difference in how the ages of node-calibration-fossils reflect the actual age of their respective node. Further, node-calibration-fossils at deeper nodes tend to have lower ESFs.

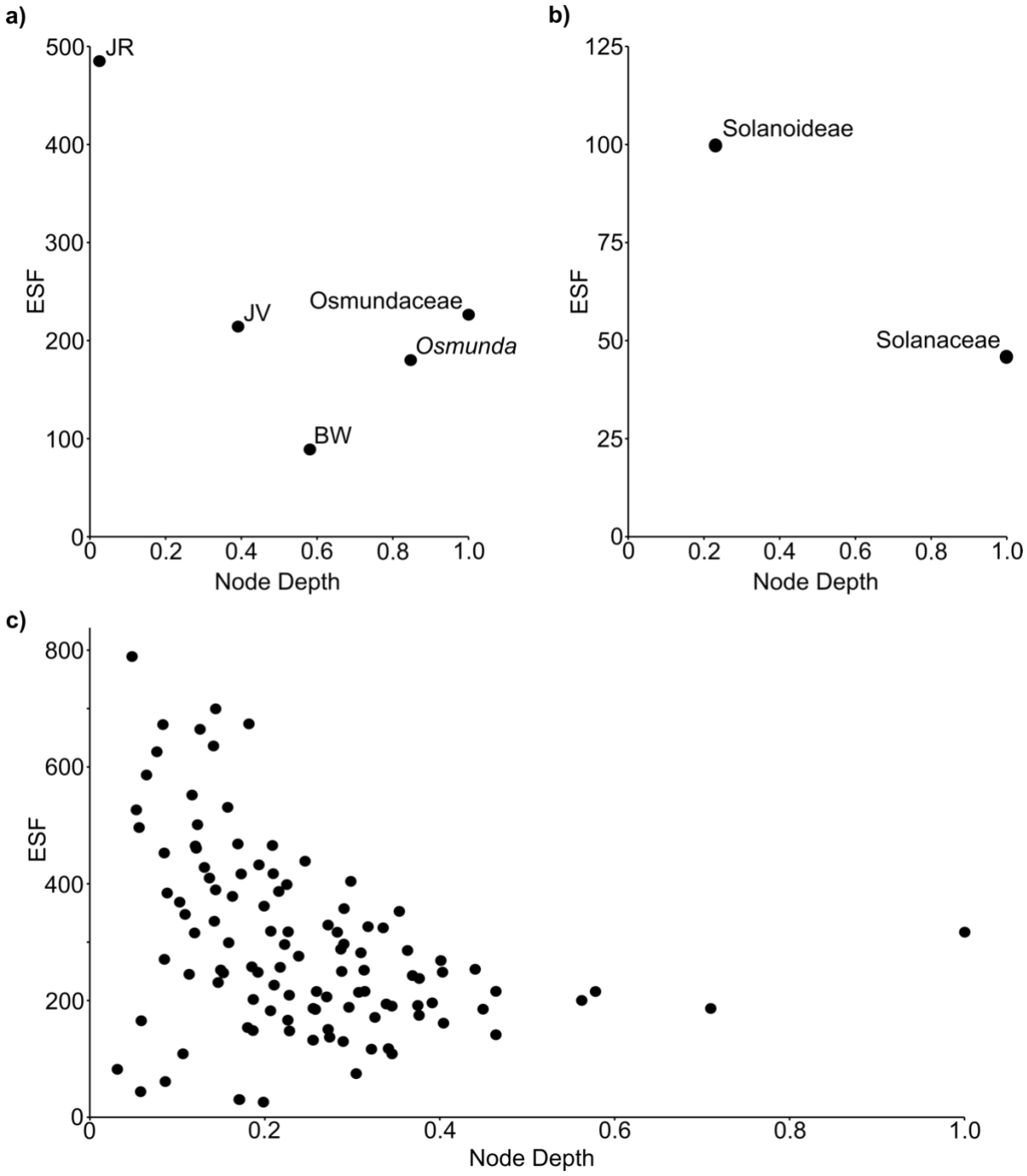

Figure 3.6. The distribution of empirical scaling factors (ESFs) for sets of node-calibration-fossils previously implemented in a) Osmundaceae, b) CS, and c) Spermatophyta. The ESF provides an approximation of how node-calibration-fossils reflect the actual age of their respective clades. Comparison of ESFs for a set of node-calibration-fossils therefore provides a basis to determine the extent to which they reflect the relative ages of different clades. In all three clades, the range of ESFs suggests sets of node-calibration-fossils are unlikely to reflect the relative ages of different clades.

The fossil recovery rate (ψ)

Osmundaceae

Large differences in ψ were estimated within Osmundaceae (Fig. 3.7a). The highest estimate of ψ is for JV (excluding JR), with 18.4 fossils per-unit-of-time. The lowest estimate of ψ is for BW, with 0.7 fossils per-unit-of-time.

CS

Large differences in ψ were also estimated within CS (Fig. 3.7b). The estimate of ψ for Solanaceae (including the stem lineage but excluding Solanoideae) is 2.1 fossils per-unit-of-time, the estimate of ψ for Solanoideae (including the stem lineage) is 11.7 fossils per-unit-of-time, and the estimate of ψ for Convolvulaceae (including the stem lineage) is 0.0 fossils per-unit-of-time.

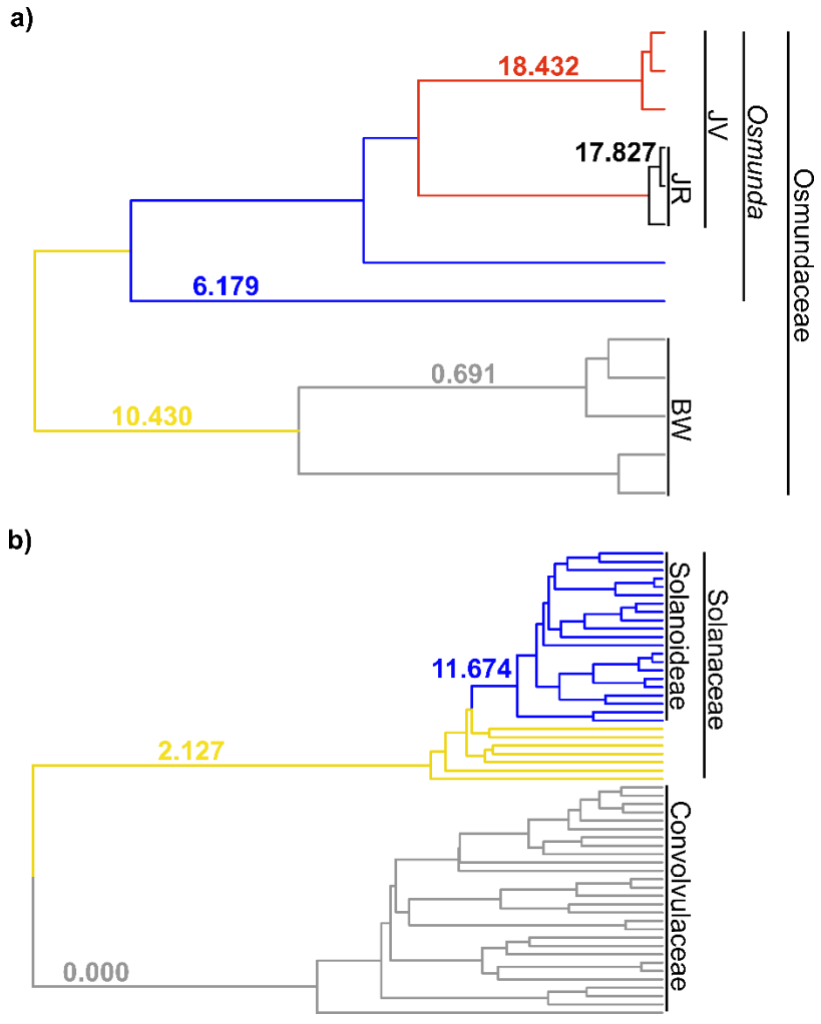

Figure 3.7: Inferred ψ for different segments of the evolutionary history of a) Osmundaceae, and b) CS. Trees shown are ultrametric phylogenies inferred in treePL with the optimum smoothing value. The FBDP assumes a constant ψ, whilst the analysis presented here suggests ψ is highly heterogeneous in Osmundaceae and CS.

Inferring divergence times with different calibration strategies and molecular clock strategies

Mean posterior estimates (MPEs)

For node calibration strategies with non-uniform calibration densities, NUP leads to younger MPEs than NUI. With NUI, there is typically a 1.15 to 1.40-fold increase in MPEs relative to NUP (Table 3.5 and 3.6, Fig. 3.8 and 3.9). For node calibration strategies with uniform calibration densities, UP leads to younger MPEs than UI. With UI, there is typically a 1.15 to 1.40-fold increase in MPEs relative to UP (Table 3.5 and 3.6, Fig. 3.8 and 3.9). The ranked order of MPEs for different calibration strategies is also sensitive to the molecular clock strategy (Table 3.5 and 3.6, Fig. 3.8 and 3.9), and the ranked order at different nodes varies most with no molecular data, less with a relaxed clock, and least with a strict clock (Fig. 3.8 and 3.9).

For the FBDP, MPEs are intermediate between the oldest and youngest MPEs inferred with node calibration strategies (Table 3.5, Fig. 3.8). On average, there is a 1.05-fold increase in MPEs with a relaxed clock relative to a strict clock, and the difference is larger at younger nodes (Table 3.5, Fig. 3.8).

Table 3.5. Mean MPEs for Osmundaceae, expressed as a percentage difference relative to NUP with a strict clock

	NUP	**NUI**	**UP**	**UI**	**FBDP**
Strict Clock	--	+30.0	+36.7	+93.8	+31.1
Relaxed Clock	-8.3	+26.1	+38.4	+88.2	+38.0
No Molecular Data	+78.5	+123.6	+443.7	+521.5	--

Table 3.6. Mean MPEs for CS, expressed as a percentage difference relative to NUP with a strict clock

	NUP	**NUI**	**UP**	**UI**
Strict Clock	--	+16.7	-11.3	+12.1
Relaxed Clock	-4.8	+10.5	+10.9	+33.0
No Molecular Data	+82.3	+142.3	+237.8	+349.4

Difference between the MPE and the oldest fossil within a clade

For node calibration strategies with non-uniform calibration densities, NUP leads to smaller differences between the age of the oldest fossil for a clade and the MPE for the clade, compared to

NUI. With NUI, there is typically a 2 to 15-fold increase in the difference relative to NUP, although in some cases the increase is much greater (Table 3.7-3.12). For node calibration strategies with uniform calibration densities, UP leads to smaller differences than UI. With UI, there is typically a 1.5 to 2.5-fold increase in the difference relative to UP (Table 3.7-3.12).

The difference is also sensitive to the molecular clock strategy. For non-uniform calibration densities (NUP and NUI), the difference becomes progressively smaller with a relaxed clock and no molecular data, compared to a strict clock (Table 3.7-3.12). This is most clearly the case for NUP, and especially if the node-calibration-fossil has a low ESF, such as the Solanaceae stem node fossil (Table 3.2, 3.4, 3.7-3.12). For uniform calibration densities (UP and UI), the effect of different molecular clock strategies is more complex. For older node-calibration-fossils (especially with low ESFs), the difference becomes progressively smaller with a relaxed clock and no molecular data compared to a strict clock (Table 3.2, 3.4, 3.7-3.12). Further, the decrease in difference is greater for UI than UP, but the overall scale of the differences are larger for UI than UP, especially with a strict clock (Table 3.7-3.12). By contrast, for uniform calibration densities and younger node-calibration-fossils (especially with low ESFs), the difference becomes somewhat larger with a relaxed clock and no molecular data compared to a strict clock (Table 3.2, 3.4, 3.7-3.12).

For the FBDP, the differences are intermediate between the smallest and largest differences inferred with node calibration strategies (Table 3.7 and 3.8). There were only moderate changes in the difference between the strict clock and the relaxed clock (Table 3.7 and 3.8). With a relaxed clock, the differences were smaller for clades/grades with a lower inferred ψ such as *Osmunda* (excluding JR and JV) and BW, and larger for clades/grades with a higher inferred ψ such as JR (Table 3.7 and 3.8, Fig. 3.7a).

Table 3.7. Percentage difference in age between the MPE and oldest fossil for different clades within Osmundaceae, when divergence times are inferred with a strict clock

Clade	Percentage difference in age for different calibration strategies				
	NUP	**NUI**	**UP**	**UI**	**FBDP**
Osmundaceae	0.4	18.4	23.3	81.0	9.9
Osmunda	37.9	72.5	79.9	163.8	60.0
JV	0.9	42.7	50.5	118.3	36.6

BW	60.6	143.3	157.7	277.4	121.8
JR	2.3	8.4	7.6	16.2	31.3
Mean	**20.4**	**57.1**	**63.8**	**131.3**	**51.8**

Table 3.8. Percentage difference in age between the MPE and oldest fossil for different clades within Osmundaceae, when divergence times are inferred with a relaxed clock

Clade	Percentage difference in age for different calibration strategies				
	NUP	**NUI**	**UP**	**UI**	**FBDP**
Osmundaceae	0.8	13.4	19.6	71.0	9.3
Osmunda	5.3	55.4	66.3	136.9	53.0
JV	0.6	21.8	35.0	86.4	34.1
BW	4.7	114.9	160.0	270.7	107.2
JR	3.0	11.8	13.9	23.5	56.2
Mean	**2.9**	**43.5**	**58.9**	**117.7**	**52.0**

Table 3.9. Percentage difference in age between the MPE and oldest fossil for different clades within Osmundaceae, when divergence times are inferred with no molecular data

Clade	**Percentage difference in age for different calibration strategies**			
	NUP	**NUI**	**UP**	**UI**
Osmundaceae	1.2	29.4	16.5	56.7
Osmunda	2.1	21.1	43.4	77.2
JV	0.6	20.2	66.9	91.9
JR	2.7	35.8	238.6	318.4
BW	4.6	40.5	598.3	685.4
Mean	**2.2**	**29.4**	**192.8**	**245.9**

Table 3.10. Percentage difference in age between the MPE and oldest fossil for different clades within CS, when divergence times are inferred with a strict clock

Clade	Percentage difference in age for different calibration strategies			
	NUP	**NUI**	**UP**	**UI**
Solanaceae (stem)	120.0	157.9	85.7	147.8
Solanoideae (stem)	3.9	19.4	1.9	14.6
Mean	62.0	88.6	43.8	81.2

Table 3.11. Percentage difference in age between the MPE and oldest fossil for different clades within CS, when divergence times are inferred with a relaxed clock

Node	Percentage difference in age for different calibration strategies			
	NUP	**NUI**	**UP**	**UI**
Solanaceae (stem)	34.9	62.9	64.1	99.0
Solanoideae (stem)	3.5	15.0	15.2	36.7
Mean	19.2	39.0	39.7	67.8

Table 3.12. Percentage difference in age between the MPE and oldest fossil for different clades within CS, when divergence times are inferred with no molecular data

Node	Percentage difference in age for different calibration strategies			
	NUP	**NUI**	**UP**	**UI**
Solanaceae (stem)	3.1	20.9	44.4	91.5
Solanoideae (stem)	37.5	91.6	154.1	229.6
Mean	20.3	56.2	99.2	160.5

95% HPD widths

For node calibration strategies with non-uniform calibration densities, NUP leads to narrower 95% HDPs than NUI. With NUI, there was typically a 1.1 to 1.5-fold increase in the 95% HPD width relative to NUP (Table 3.13 and 3.14). For node calibration strategies with uniform calibration densities, UP leads to narrower 95% HPDs than UI. With UI, there was typically a 1.15 to 1.35-fold increase in the 95% HPD width relative to UP (Table 3.13 and 3.14). In both datasets, 95% HPDs tend to be narrowest with the strict clock, wider with the relaxed clock (around 1.3-1.4 times wider than the strict clock), and widest with no molecular data (often over twice as wide as the strict clock) (Table 3.13 and 3.14). For the FBDP, there was a 1.3-fold increase in the 95% HPD width with a relaxed clock relative to a strict clock (Table 3.13).

Table 3.13. Mean 95% HPD widths as a percentage of MPE for Osmundaceae

	NUP	**NUI**	**UP**	**UI**	**FBDP**
Strict Clock	53.4	72.8	61.6	73.0	56.6
Relaxed Clock	63.8	82.7	84.2	98.7	73.2
No Data	121.5	152.6	173.8	199.6	--

Table 3.14. Mean 95% HPD widths as a percentage of MPE for CS

	NUP	**NUI**	**UP**	**UI**
Strict Clock	51.1	79.6	45.6	56.8
Relaxed Clock	69.5	78.3	72.4	98.7
No Data	132.4	143.9	141.8	161.3

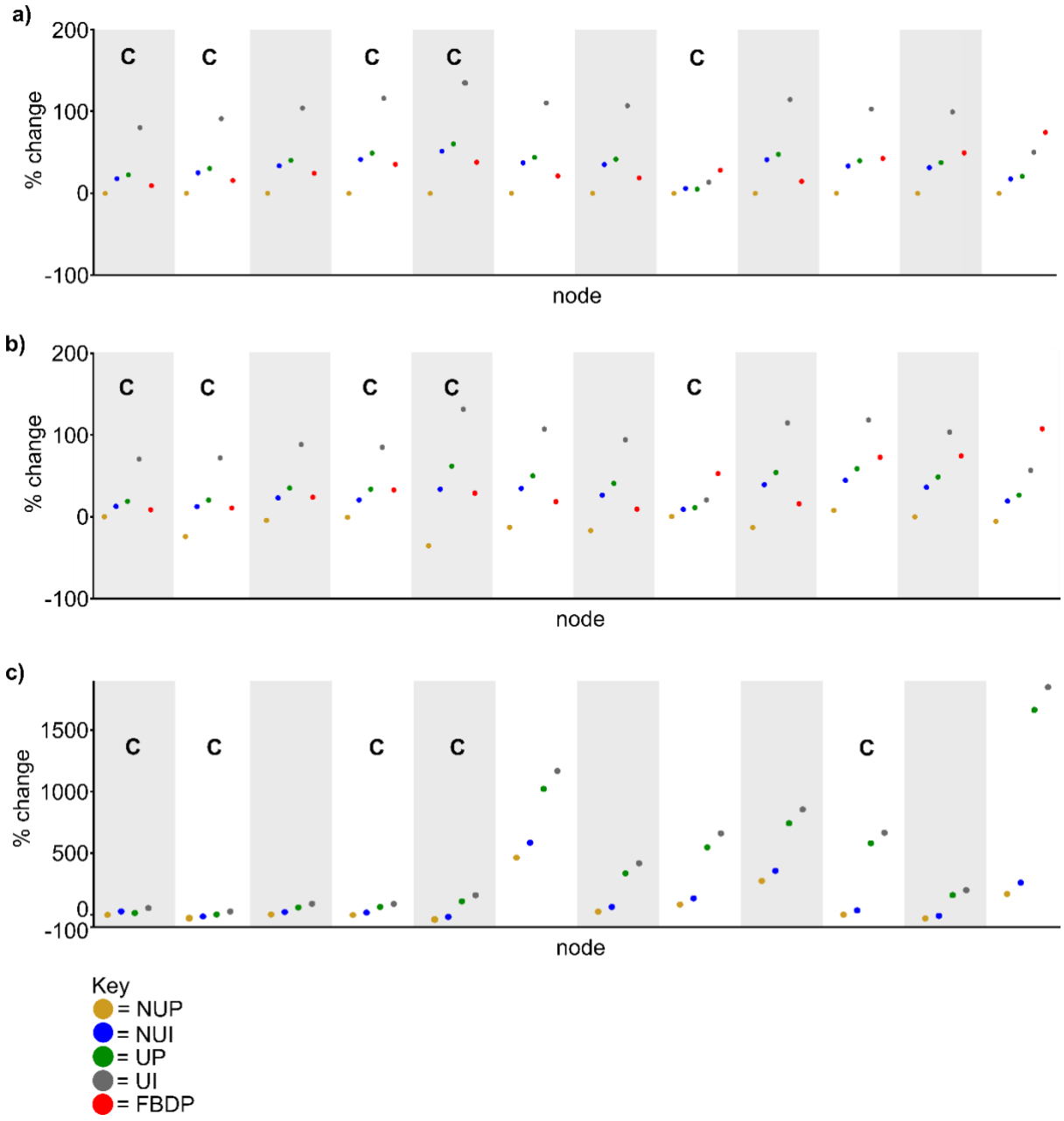

Figure 3.8. Inferred divergence times in Osmundaceae with different calibration strategies and different molecular clock strategies. Plotted points refer to the percentage change of the MPE for a given node and inference method, relative to the MPE for the node with NUP and a strict clock. Different coloured points refer to different calibration strategies (see key). In a) divergence times are inferred with a strict clock, in b) divergence times are inferred with a relaxed clock, and in c) divergence times inferred with no molecular data. Each grey or white stripe refers to a different node – these are ordered by depth in an ultrametric phylogeny of the same taxa (see methods).

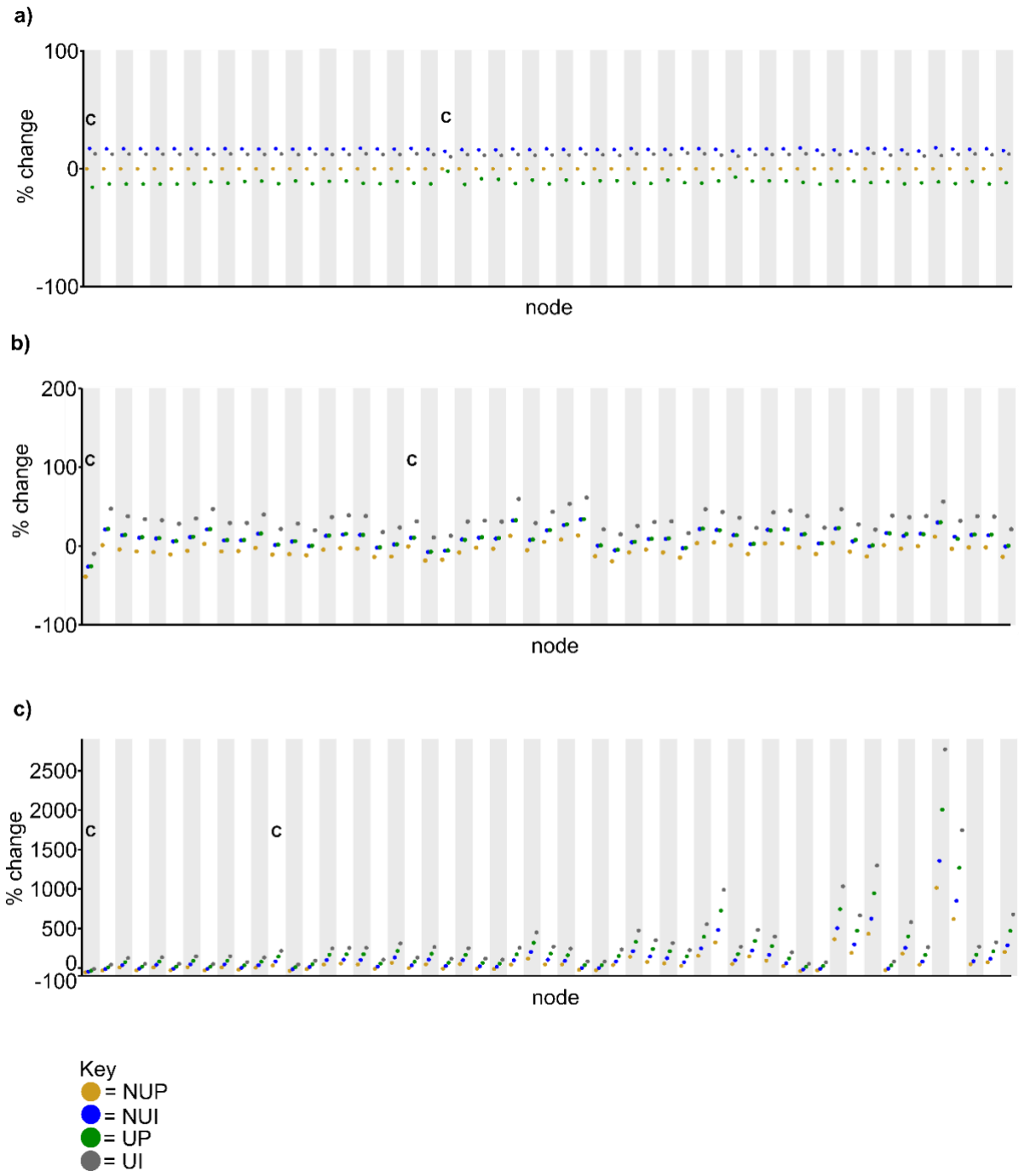

Figure 3.9. Inferred divergence times in CS with different calibration strategies and different molecular clock strategies. Plotted points refer to the percentage change of the MPE for a given node and inference method, relative to the MPE for the node with NUP and a strict clock. Different coloured points refer to different calibration strategies (see key). In a) divergence times are inferred with a strict clock, in b) divergence times are inferred with a relaxed clock, and in c) divergence times inferred with no molecular data. Each grey or white stripe refers to a different node – these are ordered by depth in an ultrametric phylogeny of the same taxa (see methods).

DISCUSSION

Multiple node-calibration-fossils are unlikely to reflect the relative ages of different clades

There is considerable uncertainty in how the ages of node-calibration-fossils reflect the actual ages of their respective clades in both Osmundaceae and CS. Multiple node-calibration-fossils are therefore unlikely to reflect the relative ages of different clades. This conclusion is based on the inferred confidence intervals for the ages of clades within Osmundaceae and CS. The timespan covered by these confidence intervals differ profoundly, and cover timescales that are 10s to 100s of millions-of-years older than the node-calibration-fossil (Fig. 3.5a and Fig. 3.5b).

Given the upper limits of the confidence intervals are vastly older than the 95th percentiles of calibration densities from Grimm et al. (2015) and Särkinen et al. (2013), the confidence intervals also highlight that the assumptions of previous studies are unrealistic (Fig. 3.5a and 3.5b). Even if the confidence intervals are ignored, Figures 3.5a and 3.5b illustrate problems with the assumptions of Grimm et al. (2015) and Särkinen et al. (2013). Gaps in the fossil records of Osmundaceae and CS are too extensive to make confident statements about clade ages in relation to the ages of node-calibration-fossils, as was done in these studies. In this context, the discovery of plant fossils that are vastly older than previously known fossils and previous divergence time estimates (Wilf and Escapa 2015) – including a 52 million-year-old Solanoideae fossil not included in this study (Wilf et al. 2017) – is unsurprising.

The conclusion that multiple node-calibration-fossils are unlikely to reflect the relative ages of different clades is corroborated by the ESFs that were calculated for node-calibration-fossils in Osmundaceae, CS, and Spermatophyta (Fig. 3.6a, b, and c). These suggest considerable differences in how the ages of node-calibration-fossils reflect the actual ages of their respective clades (more than 5-fold differences in Osmundaceae, more than 2-fold differences in CS, and more than 29-fold differences in Spermatophyta).

Analyses of ESFs are confounded by among-branch-substitution-rate-variation. Conclusions about fossil calibrations in the Spermatophyta dataset – which are based solely on ESFs – are

therefore less robust. However, variation in ESFs within Spermatophyta is of a larger scale than Osmundaceae and CS. Given that confidence intervals within Osmundaceae and CS corroborate the argument that multiple node-calibration-fossils are unlikely to reflect the relative ages of different clades, it cannot be assumed with confidence that this is not also the case in Spermatophyta.

Variation in ψ

Recent discussion has focussed on the value of the FBDP for integrating fossils with molecular phylogenies to estimate divergence times. It is often suggested that a strength of the FBDP is that unlike node calibration methods, it explicitly models fossil occurrences and extracts temporal information from the entire fossil record (Heath 2014; Grimm et al. 2015; Lee and Palci 2015; Renner et al. 2016; Gavryushkina et al. 2017). In order to do this, ψ is inferred. However, an underlying assumption is that ψ is constant. The analyses of Osmundaceae and CS presented here indicate that contrary to this assumption, ψ is highly heterogeneous (Fig. 3.7a and b). This indicates that the manner by which the FBDP extracts temporal information from the fossil record is fundamentally undermined. This is likely to be a problem across much of the tree of life, given that analyses of empirical datasets often indicate that a constant ψ is an unreasonable assumption to make about the fossil record (Raup 1972; Signor and Lipps 1982; Smith 2001; Smith and McGowan 2005).

Similarly to ESFs, estimates of ψ are confounded by among-branch-substitution-rate-variation. Regardless, Figure 3.4 (which summarises the phylogenetic distribution of fossils) highlights that large differences in ψ are still highly likely. For example, within Osmundaceae there is a 27-fold difference between the number of fossils in two sister clades (Fig. 3.4a).

Comparing divergence time estimates with different methods

Comparing MPEs for node calibration strategies

When comparing node calibration strategies that use either non-uniform calibration densities (NUP and NUI) or uniform calibration densities (UP and UI), the strategies that assume actual clade ages are closer to the ages of node-calibration-fossils (NUP or UP) lead to younger MPEs than the strategies that assume actual clade ages may be considerably older (NUI or UI) (Table 3.5 and 3.6;

Fig. 3.8 and 3.9). This pattern occurs regardless of the molecular clock strategy, thus highlighting the importance of the calibration strategy in all contexts. This result, which demonstrates the general importance of the calibration strategy, is consistent with many previous studies (Donoghue and Benton 2007; Warnock et al. 2011, 2015; Magallón et al. 2013; Betts et al. 2018).

However, these analyses also highlight the sensitivity of MPEs to the molecular clock strategy in two key ways. First, the ranked order of MPEs with different calibration strategies at each node becomes progressively more varied with a relaxed clock and no molecular data, compared to a strict clock. Second, MPEs inferred with a relaxed clock tend to be intermediate between a strict clock and no molecular data (Table 3.5 and 3.6: Fig. 3.8 and 3.9).

For both of these observations about molecular clock strategies, the relaxed clock is intermediate between the strict clock and no molecular data. This pattern can be explained by comparing the attributes of the three molecular clock strategies: the strict clock does not allow among-branch-substitution-rate-variation; the relaxed clock allows some among-branch-substitution-rate-variation but its magnitude and nature is constrained by the model used (in these experiments a UCLN relaxed clock); and analyses with no molecular data are equivalent to assuming that rates can vary among branches by an infinite magnitude. Therefore, when multiple fossil calibrations are used with a strict clock, they do not influence divergence time estimates by affecting estimates of branch-specific rates. By contrast, when multiple fossil calibrations are used with a relaxed clock, they can influence divergence time estimates by affecting estimates of branch-specific rates, but in a manner that is constrained by the model of rate variation. Meanwhile, when multiple fossil calibrations are used with no molecular data, they can influence divergence time estimates in a manner that is completely unconstrained by assumptions about rates. In this context, it is unsurprising that analyses with a relaxed clock are intermediate between a strict clock and no molecular data.

Difference between the MPE and age of the node-calibration-fossil (the oldest fossil within a clade)

When comparing node calibration strategies that use either non-uniform calibration densities (NUP and NUI) or uniform calibration densities (UP and UI), the strategies that assume actual clade

ages are more similar to the ages of node-calibration-fossils (NUP or UP) lead to smaller differences between MPEs and the ages of the node-calibration-fossils than strategies that assume actual clade ages may be considerably older (NUI or UI) (Table 3.7-3.12). This pattern occurs regardless of the molecular clock strategy, thus highlighting the importance of the calibration strategy in all contexts.

However, the relationship between MPEs and ages of node-calibration-fossils is strongly influenced by the molecular clock strategy, and interactions between specific calibration strategies and molecular clock strategies. For example, with NUP and NUI, the strict clock results in the largest differences (Table 3.7 and 3.10), followed by the relaxed clock (Table 3.8 and 3.11), followed by no molecular data (Table 3.9 and 3.12).

This pattern can be explained within the same framework as for the effects of different molecular clock strategies on MPEs. First, consider the strict clock – which infers the same rate for every branch, and NUP and NUI – which specify small but non-zero probabilities that clades are markedly older than node-calibration-fossils (Fig. 3.3a and b). In this context, MPEs that are vastly older than node-calibration-fossils can be the most feasible divergence time estimates. By contrast, relaxed clocks infer different rates for each branch, enabling MPEs to correspond more closely to the specified calibration density. This results in smaller differences between MPEs and the ages of node-calibration-fossils, especially for NUP, which assumes the actual clade age is very similar to the age of the node-calibration-fossil. Meanwhile, with no molecular data, assumptions about rates confer no constraints on MPEs. As such, MPEs correspond exactly (or almost exactly after accounting for the effect of the tree prior) to the specified calibration density. This result is consistent with that of Brown and Smith (2018), who showed that with a relaxed clock, posterior age estimates at calibrated nodes are often very similar to age estimates inferred with no molecular data.

In relation to this pattern, in CS with a strict clock and NUP or NUI, the difference is far larger at the Solanaceae stem node than the Solanoideae stem node (Table 3.10). The difference at the Solanaceae stem node then becomes progressively smaller with a relaxed clock (Table 3.11) and no molecular data (Table 3.12). This is likely to reflect the fact that the ESF for the node-calibration-fossil at the Solanoideae stem node is higher than at the Solanaceae stem node (Fig. 3.6b, Table 3.4).

With a strict clock, the higher ESF at the Solanoideae stem node means the entire phylogeny is calibrated to a timescale that is far older than the node-calibration-fossil at the Solanaceae stem node. By contrast, with a relaxed clock or no molecular data there is no assumption of rate constancy, meaning the calibration at the Solanoideae stem node does not influence MPEs throughout the phylogeny in as consistent a manner. The MPE for the Solanaceae stem node therefore corresponds more closely to the calibration density specified by NUP or NUI.

With UP and UI the results are more complicated because of three characteristics of these calibration strategies. First, UP and UI do not make an explicit statement about the probable clade age in relation to the node-calibration-fossil, other than it lies somewhere between the minimum and maximum. Second, UP and UI implement a hard-maximum constraint, the clade can be no older than this maximum. Third, the same maximum is used for several different nodes because robust maximum constraints are difficult to justify (Marshall 2008; Warnock et al. 2015, Betts et al. 2018; Morris et al. 2018). For nodes with younger node-calibration-fossils, the range between the minimum and maximum is therefore often larger.

These differences have several implications. First, a relaxed clock or no molecular data does not cause MPEs to correspond to the ages of node-calibration-fossils in as predictable a manner as the case for NUP and NUI (Table 3.7-3.12). This is because UP and UI do not make an explicit statement about probable clade ages in relation to node-calibration-fossils, and because there are different sized intervals between minimum and maximum constraints at each node.

With UI, and for node-calibration-fossils with lower ESFs, there are larger differences between the ages of node-calibration-fossils and MPEs (Table 3.2, 3.4, 3.7-3.12). For older node-calibration-fossils, these differences become progressively smaller with a relaxed clock and no molecular data compared to a strict clock. This can be explained according to the same framework as set out for NUP and NUI in CS. Specifically, relaxing the assumption of rate constancy means that node-calibration-fossils with higher ESFs do not calibrate the entire phylogeny to a timescale that is older than node-calibration-fossils with lower ESFs. However, for younger node-calibration-fossils, these differences do not become progressively smaller with a relaxed clock and no molecular data

compared to a strict clock. This is likely to be because of the larger range between the minimum and maximum constraints at the nodes that these fossils calibrate, and the ambiguity about the clade age within this range.

With UP and for older node-calibration-fossils with lower ESFs, there are less noticeable differences between molecular clock strategies (Table 3.2, 3.4, 3.7-3.12). The younger hard maximum prevents MPEs from being markedly older than the node-calibration-fossil, regardless of molecular clock strategy or ESF. For UP and younger node-calibration-fossils with lower ESFs, different molecular clock strategies also have a limited effect, as was also the case with UI. However, the magnitude of the differences is smaller than for UI.

95% HPD widths

When comparing node calibration strategies that use either non-uniform calibration densities (NUP and NUI) or uniform calibration densities (UP and UI), the strategies that assume actual clade ages are more similar to the ages of node-calibration-fossils (NUP or UP) lead to narrower 95% HPDs than strategies that assume actual clade ages may be considerably older (NUI or UI) (Table 3.13 and 3.14). This pattern occurs regardless of the molecular clock strategy, thus highlighting the importance of the calibration strategy in all contexts. This result, which demonstrates the general importance of the calibration strategy to the 95% HPD width, is consistent with several previous studies (Warnock et al 2012; Magallón et al. 2013; Betts et al. 2018).

In all cases, 95% HPDs are narrowest with a strict clock, and progressively wider with a relaxed clock and no molecular data (Table 3.13 and 3.14). This can be explained within the same framework as the previous discussion of why the relaxed clock is intermediate between the strict clock and no molecular data. With weaker assumptions about rates for specific branches (from the strict clock, to relaxed clock, to no molecular data), there is a larger range of probable divergence time estimates, and hence wider 95% HPDs.

The FBDP

Analyses with the FBDP allow this newly developed method to be compared to conventional node calibration methods. Although these analyses are more limited in scope, several similar characteristics between these two methods are identified. First, 95% HPD widths and MPEs are highly sensitive to molecular clock strategy (Table 3.5 and 3.13, Figure 3.8a and b). Although the details of this relationship are complex for MPEs, and depend on the phylogenetic depth at which a comparison is made (Fig. 3.8a and b), for 95% HPD widths the picture is simple. 95% HPDs are broader with a relaxed clock (Table 3.13). This is consistent with node calibration methods, and the same explanation is proposed as previously, whereby weaker assumptions about rates lead to a larger range of probable divergence time estimates.

A second key similarity is that with a relaxed clock, inferred divergence times are more consistent with assumptions about the fossil record. With node calibration methods, this conclusion is based on the fact that with a relaxed clock, inferred ages for calibrated nodes correspond more closely to the specified calibration density (Table 3.7-3.12). For the FBDP, this conclusion is based on the fact that with a relaxed clock, in clades with a lower inferred ψ, divergence time estimates tend to be closer to the age of the oldest fossil (and are therefore younger), and in clades with a higher inferred ψ, divergence time estimates tend to be less close to the age of the oldest fossil (and are therefore older) (Table 3.7 and 3.8, Fig. 3.7a). This observation is likely to reflect the assumption that ψ is constant. With a relaxed clock, the inference of different rates for each branch means that younger ages are inferred for clades with a lower inferred ψ, therefore "smoothing" differences in ψ among clades. By contrast, the opposite occurs in clades with a higher inferred ψ. There are nonetheless, some exceptions to this pattern. These are likely to be because the inferred ages of different clades in a phylogeny are not independent of each other.

Relaxed clocks and multiple fossil calibrations do not reliably lead to more accurate divergence time estimates

The development of methods for integrating multiple fossil calibrations with relaxed clocks has led to increased agreement between the ages of known fossils and inferred divergence times. This is often interpreted as scientific "progress", with methodological developments enabling integration of temporal evidence from the fossil record with molecular sequence data to construct a more realistic timeframe for the tree of life (Donoghue and Benton 2007; Grimm et al. 2015; Hughes et al. 2015; Magallón et al. 2015; Donoghue and Yang 2016; Betts et al. 2018; Morris et al. 2018). Here, I show that this positive characterisation is undermined in two ways. First, the nature of the fossil record is inconsistent with the assumptions of current methods. Second, methods that integrate multiple fossil calibrations with relaxed clocks are highly sensitive to these assumptions.

For node calibration methods, I conclude that in the context of the plant clades analysed in this study, which incorporate a range of phylogenetic scales and for which the fossil records have a range of different characteristics, the nature of the fossil record is inconsistent with the assumptions of current methods because analyses highlight that sets of node-calibration-fossils are unlikely to reflect the relative ages of different clades. For the FBDP, I conclude that the nature of the fossil record is inconsistent with methodological assumptions based on analyses which highlight that ψ is highly heterogeneous.

Subsequently, comparisons of inferred divergence times when using multiple fossil calibrations with different methods highlight the sensitivity of methods to underlying assumptions. I show that divergence time estimates are highly sensitive to the calibration strategy, regardless of molecular clock strategy. I also show that divergence time estimates are sensitive to the molecular clock strategy, with context specific interactions between calibration strategies and molecular clock strategies.

Further, an important general pattern is that analyses with relaxed clocks are intermediate between strict clocks and no molecular data. This pattern is important because it indicates that

analyses which use multiple fossil calibrations with relaxed clocks are often unlikely to provide novel or robust insights. Instead, the molecular sequence data does not provide direct information about rate and time. Therefore, in analyses with relaxed clocks whereby different rates are inferred on individual branches, divergence time estimates will be sensitive to implemented fossil calibrations and the constraints of the relaxed clock model, whilst when no molecular data is analysed in divergence time analyses, divergence time estimates will solely be influenced by implemented fossil calibrations (and potentially also the birth-death tree prior).

The purpose of this study is not to directly challenge the results of previous studies. Instead, by showing that current methods are highly sensitive to a set of assumptions that are often violated, I suggest that methods that use multiple fossil calibrations with relaxed clocks often do not produce robust inferences. As such, it may often be the case that analyses that use multiple fossil calibrations with relaxed clocks do not produce meaningful inferences that add significantly to those that can be made by simply analysing the fossil record. I therefore challenge a common assertion that highly integrated analyses that use multiple fossil calibrations with relaxed clocks inherently lead to more meaningful, robust, and novel inferences (Heath et al. 2014; Grimm et al. 2015; Lee and Palci 2015; Magallón et al. 2015; Renner et al 2016; Yang and Donoghue 2016; Betts et al. 2018; Morris et al. 2018).

Implications for macroevolutionary studies

The findings presented in this paper have concerning implications for macroevolutionary studies, because many macroevolutionary studies are heavily dependent on accurate time-calibrated phylogenies. This raises important questions about how time-calibrated phylogenies should best be inferred and interpreted. One potential option for node calibration methods, could be to implement calibration densities that are markedly broader than in many previous studies, or maximum age constraints that are markedly older. This approach was followed in the NUI and UI calibration strategies discussed here. However, such an approach makes extremely weak assumptions about times such that it is difficult to meaningfully implement different molecular clock strategies. Further, 95%

HPDs (or equivalent confidence intervals) would be extremely broad. Using inferred time-calibrated phylogenies as a basis for hypothesis testing would therefore likely become highly problematic.

Alternatively, a potential option for the FBDP could be to relax the assumption that ψ is constant when inferring divergence times. Although extensions to the FBDP are continually being developed (for example Stadler et al. 2018), no method for estimating divergence times which enables ψ to vary has been implemented. Further, the prospects of developing a method of divergence time estimation which enables ψ to vary are likely to be limited because it would cause fundamental issues with model identifiability. Any such method would also likely be undermined by the continued assumption that λ and μ are constant in the underlying branching process.

Rather than hoping for methodological developments which may be theoretically impossible, a useful approach for future studies may be to change the way in which time-calibrated phylogenies are interpreted. Instead of using time-calibrated phylogenies as a basis to construct general narratives for the evolution of a particular group (for example Lagomarsino et al. 2016; Cardillo et al. 2017; Folk et al. 2019), a more robust approach may be to test specific evolutionary hypotheses. With such an approach, analyses can be constructed that account for the biases inherent in different methods and provide a robust test of a particular hypothesis. This approach was followed in a recent study of the sweet potato, where calibration strategies biased divergence time estimates to younger ages to provide the most robust test of whether the sweet potato evolved in pre-human times (Munoz-Rodriguez et al. 2018, 2019, Carruthers et al. 2020). In this context, no assumptions are required about whether or not a given time-calibrated phylogeny is most accurate, a question which is often impossible to answer.

4. TEMPORAL DYNAMICS OF EVOLUTIONARY DIVERSIFICATION IN *IPOMOEA*

ABSTRACT

Molecular phylogenies are used as a basis for making inferences about macroevolutionary history. However, a robust phylogeny does not contain the information that is necessary to make many of these inferences. Complex methodologies that incorporate important assumptions about the nature of evolutionary history are therefore required. Here, I explore the implications of these assumptions for making inferences about the macroevolutionary history of *Ipomoea* – a large pantropical genus of flowering plants that contains the economically important crop sweet potato (*Ipomoea batatas*). I focus on assumptions that underlie inferences of divergence times and diversification parameters, two of the most fundamental variables in macroevolutionary research. Specifically, I use a series of novel approaches to explore the implications of these assumptions for inferring the age of *Ipomoea*, the ages of major clades within *Ipomoea*, whether there are significant differences in diversification parameters among clades within *Ipomoea*, and whether the storage root of *I. batatas* evolved in pre-human times. I show that inferring an age estimate for *Ipomoea* and major clades within *Ipomoea* is highly problematic. Inferred divergence times are sensitive to uncertain fossil calibrations and differing assumptions about among-branch-substitution-rate-variation. Despite this uncertainty, robust inferences can still be made about patterns of net diversification rate variation within *Ipomoea*, and that the storage root of *I. batatas* evolved in pre-human times. Taken together, this study presents novel and generalizable insights into the implications of methodological assumptions for making inferences about macroevolutionary history. Further, by presenting novel findings relating to the temporal dynamics of evolution in *Ipomoea*, as well as more specifically to *I. batatas*, this study makes a valuable contribution to our understanding of tropical plant evolution, and the evolutionary context in which economically important crops evolve.

INTRODUCTION

Molecular phylogenies are of central importance to macroevolutionary research**.** Our understanding of evolutionary radiations, and more generally, the factors that affect when, how, and why diversity accumulates in different clades, is underpinned by the interpretation of molecular phylogenies in an evolutionary context (Baldwin and Sanderson 1998; Hughes and Eastwood 2006; Rabosky et al. 2008; Givnish et al. 2009; Koenan et al. 2013; Donoghue and Sanderson 2015; Nevado et al. 2016; Rabosky et al. 2018). So too is our understanding of when and how the Earth's major biomes were assembled (Pennington et al. 2009; Simon et al. 2009; Govindarajulu et al. 2011; Särkinen et al. 2012; Hughes et al. 2013; Pennington and Lavin 2015; Dexter et al. 2017) and the historical processes that have shaped distribution patterns such as long-distance-dispersal and geological vicariance (Lavin et al. 2004; Donoghue 2008; Crisp et al. 2009; Muñoz-Rodríguez et al. 2018). Because of their importance to macroevolutionary research, the rapid increase in availability of robust molecular phylogenies for many groups has enabled macroevolutionary questions to be investigated in greater detail and a greater variety of different contexts.

There are fundamental differences between a robust molecular phylogeny and the evolutionary process that it is used as a basis for making inferences about. These differences mean that even if a robust phylogeny is known for a group, parameters of interest in macroevolutionary research, such as divergence times and diversification parameters, can be inferred with significant error (Sanderson and Doyle 2001; Britton 2005; Rabosky 2010; Wilf and Escapa 2015; Moore et al. 2016; Carruthers et al. 2019). It is crucial to account for this error, especially when using inferences as a basis for constructing narratives or explanations about when, how and why a group evolved. If error is not adequately accounted for, complex narratives may be constructed that are not supported by available evidence (Wilf and Escapa 2015).

Here, I use the recently developed phylogenetic framework for *Ipomoea* (Muñoz-Rodríguez et al. 2018, 2019) to investigate the temporal dynamics of its evolutionary diversification. *Ipomoea* is a pantropical genus of over 800 species that exhibit a range of growth forms including climbers,

trailers, erect shrubs, and trees. They also occur in a range of habitats including cerrado, dry forest, and human disturbed areas (Wood et al. 2015). Further, the genus contains the sweet potato (*I. batatas*), a crop of global economic importance (Muñoz-Rodríguez et al. 2018). Investigating the temporal dynamics of evolutionary diversification in *Ipomoea* can therefore provide critical insights into the nature of evolutionary diversification in large tropical groups, and when, how and why important crops evolve from their wild relatives.

In this study, I specifically focus on the inference of divergence times and diversification parameters. This is because these two variables are the fundamental basis upon which other aspects of macroevolutionary research depend. When inferring these two variables, I explicitly consider different sources of uncertainty. With respect to divergence times, one of the primary sources of uncertainty are fossil calibrations (Sanderson and Doyle 2001; dos Reis and Yang 2013; Magallón et al 2013; Warnock et al. 2015; Morris et al. 2018). These are likely to be a particularly important source of uncertainty for *Ipomoea*, and many other angiosperm genera, because their fossil records are often extremely fragmentary and therefore provide a very incomplete picture of the temporal range and diversity of a given taxon (Särkinen et al. 2013; Mitchell et al. 2016; Cardillo et al. 2017; Wilf et al. 2017; Folk et al 2019). It is therefore difficult to make robust interpretations of the temporal signal in the fossil record, and to use it as a basis to calibrate phylogenies. For divergence times, I also consider the effect of different assumptions about substitution rates and the magnitude of among-branch-substitution-rate-variation (Zuckerkandl and Pauling 1962, 1965; Fitch 1976; Gillespie 1991; Sanderson 1997; Thorne et al. 1998; Britton 2005; Bromham 2006; Drummond et al. 2006; Zhu et al. 2015; Carruthers et al. 2019), as well as the effect of assuming a constant rate of speciation and extinction when performing divergence time analyses in a Bayesian framework. For diversification parameters, I explore the effect of divergence time estimation error on inferred diversification parameters. I also discuss the effects of limited taxon sampling, and the effects of extinct branches that are un-sampled in a molecular phylogeny.

When exploring these sources of uncertainty in the context of *Ipomoea*, I investigate several points that relate to the timing and tempo of evolutionary diversification in the entire genus. These

include the age of *Ipomoea*, the age of major clades within *Ipomoea*, as well as diversification parameters and the extent to which they vary among clades. Given the recently developed phylogenetic framework for *Ipomoea* resolves the relationships between *I. batatas* and its closest wild relatives (Muñoz-Rodríguez et al. 2018), I also investigate questions relating to the origin of *I. batatas*. This includes when *I. batatas* and its storage root evolved, and specifically, whether it evolved in pre-human times.

MATERIALS AND METHODS

Estimating the age of Ipomoea

To estimate the age of *Ipomoea*, a time-calibrated phylogeny was inferred for Convolvulaceae, which contains *Ipomoea*, and its sister family Solanaceae, which has a richer fossil record (Särkinen et al. 2013; Wilf et al. 2017). This time-calibrated phylogeny was based on the same molecular matrix and topology as for CS in Chapter 3 (Appendix 1 and 3). This sparsely sampled dataset, within which the same number of taxa are sampled from each family, is especially appropriate for divergence time analyses focussed on estimating the age of *Ipomoea*. This is because using this dataset is likely to reduce the effect of violating the assumption of the constant rate birth-death branching process that the diversification parameters are constant – an underlying assumption of most Bayesian methods of divergence time estimation. Specifically, sampling an equal number of taxa from each sister clade will likely cause diversification parameters to be constant amongst sampled taxa at broad sales across the phylogeny, whilst sampling a small number of taxa will mean that any violation of the assumption of constant rates amongst sampled taxa will have a statistically weaker effect on divergence time estimates.

The validity of different temporal calibrations for Convolvulaceae and Solanaceae is unclear. A range of different calibration strategies were therefore used. However, in no case were any fossil calibrations used within *Ipomoea* because there are no fossils that can reliably be assigned to the genus. This includes the recently published leaf fossils of *Ipomoea meghalayensis* (Srivastava et al. 2018), which could plausibly belong to several other genera (Hawthorne pers. comm.)

In *Calibration Strategy 1*, *2* and *3*, the recently discovered fossil, *Physalis infinemundi* (Wilf et al. 2017) was used as a basis for implementing fossil calibrations. This fossil – dated at around 52 Ma – is far older than fossil calibrations previously used in Convolvulaceae and Solanaceae, and is also considerably older than previous age estimates for either of these two families (Särkinen et al. 2013; Magallón et al. 2015; Mitchell et al. 2016). Because of this, it places previous assumptions about the timing of evolution within these clades, and the relationship between fossil ages and clade ages, into considerable doubt.

In *Calibration Strategy 1*, both the Solanoideae stem node and Solanaceae stem node are calibrated with a uniform distribution (minimum = 52 Ma, max = 130 Ma). The minimum corresponds to the age of *Physalis infinemundi*, and the maximum corresponds to the time that eudicot pollen suddenly appears in the fossil record (Doyle et al 1977). This calibration strategy assumes the Solanoideae stem node is older than 52 Ma, and the Solanaceae stem node is younger than 130 Ma. Apart from this, these calibrations do not make explicit assumptions about where the most probable clade ages lie between the minimum and maximum constraints.

Calibration Strategies 2 and *3* account for the fact that 130 Ma is most probably an overestimate of the Solanaceae stem node age – at least in the context of current understanding of the temporal dynamics of plant diversification (Särkinen et al. 2013; Magallón et al. 2015; Mitchell et al. 2016). Therefore, in *Calibration Strategy 2* the root node is calibrated with a lognormal calibration density parameterised with offset = 52, m = 10.4, v =10.4, whilst in *Calibration Strategy 3* the root node is calibrated with a lognormal calibration density parameterised with offset = 52, $m = 5.2$, $v = 2.1$. In *Calibration Strategies 2* and *3*, the minimum constraint of 52 Ma is retained at the Solanoideae stem node. The effect of Calibration Strategies 2 and 3 is therefore to shift the calibration density at the root node to ages that are significantly younger than 130 Ma, with Calibration Strategy 3 shifting the distribution to younger ages than Calibration Strategy 2.

It is difficult to state with confidence which of Calibration Strategies 1, 2, or 3 is the most realistic, given that there is no evidence upon which to derive assumptions about how much younger

than 130 Ma the root node is. Comparing these three Calibration Strategies therefore provides a basis to evaluate the implications of different assumptions about the likely age of the root node.

In *Calibration Strategy 4*, the root node is calibrated with a normal distribution ($\mu = 67.34, \sigma = 9.980$). This is a secondary calibration from Magallón et al. (2015) in which 132 fossil calibrations were used throughout Spermatophyta. This calibration strategy is therefore not entirely dependent on the interpretation of fossil calibrations within Convolvulaceae and Solanaceae. Characteristics of the analysis from which Calibration Strategy 4 is derived (Magallón et al. 2015) are likely to bias estimates to younger ages. For example, this study implements an extremely young maximum constraint on the angiosperm crown age, and employs calibration densities throughout the phylogeny which assume that fossil ages are very close to clade ages. Therefore, age estimates derived from Calibration Strategy 4 were primarily used with respect to the evolution of the sweet potato. This is because biasing estimates to younger ages provided the most robust test of whether the storage root of the sweet potato evolved in pre-human times – one of the main hypotheses that I set out to examine.

RevBayes was used to infer time-calibrated phylogenies according to the four calibration strategies. In all analyses, a GTR + G + I model was used with a birth-death branching process as the tree prior. The topology was fixed to that inferred for CS in Chapter 3, and starting trees consistent with the 4 different calibration strategies were generated in treePL (Smith and O'Meara 2012). In one set of analyses, a strict clock was used where the substitution rate was assumed to be the same for every branch. In a second set of analyses, an uncorrelated lognormal (UCLN) relaxed clock (Drummond et al. 2006) was used. For each analysis, two independent runs were performed. Sufficient mixing and convergence was assessed in Tracer v1.6.0. A 25% burn-in was used prior to calculating maximum a posteriori (MAP) trees, mean posterior estimates (MPEs), and 95% highest posterior density intervals (HPDs).

Inferring the ages of major clades within Ipomoea

The mean posterior estimate (MPE) for the age of *Ipomoea* from Calibration Strategy 1 was used as a point calibration for the root node of more densely sampled phylogenies for the genus. This MPE was used because the calibration strategy from which it is derived does not make any explicit assumptions about probable clade ages in relation to minimum or maximum constraints. It can therefore act as a basis from which to discuss the implications of calibration strategies that make explicit assumptions about probable clade ages in relation to minimum or maximum constraints.

Because no further calibrations are used within *Ipomoea*, the implications of root node calibrations for *Ipomoea* derived from different calibration strategies can be easily interpreted. With different root node calibrations, the relative ages of nodes within the time-calibrated phylogeny will not change, only the absolute timescale over which the time-calibrated phylogeny occurs. This also explains why I use a point calibration for the root node, rather than a calibration density. The effects of root node calibrations that derive from different calibration strategies can easily be compared. By contrast, it would be difficult to construct a single calibration density that adequately accounts for uncertainty in root node estimates that derive from each of the calibration strategies outlined above.

For divergence time analyses within *Ipomoea*, the primary interest was to estimate the ages of major clades from Muñoz-Rodríguez et al. (2019); the Parana Clade – a diverse clade centred in central South America; the Coriaceous Sepal Clade – a diverse clade centred in the Carribean; and *Ipomoea* sect. Batatas – the clade of *I. batatas* and its closest wild relatives. A further motivation was to construct a temporal framework to provide a basis for subsequent analyses of diversification parameters.

These divergence time analyses were based on the molecular datasets and inferred phylogenies from Muñoz-Rodríguez et al. (2019). This includes a nuclear dataset of 434 single copy genes for 211 species of *Ipomoea*, and a plastome dataset of 206 species of *Ipomoea*. Within these more densely-sampled datasets, there may be significant differences in diversification parameters among clades. This violates a key assumption of Bayesian methods of divergence time estimation that

use a constant rate birth-death branching process as the tree prior. treePL was therefore used to infer divergence times – a penalised likelihood method that does not use a constant rate birth-death branching process (Smith and O'Meara 2012). In treePL, analyses were performed with the nuclear and chloroplast phylogenies from Muñoz-Rodríguez et al. (2019) as input trees. Following extensive cross-validation, a smoothing value of 0.01 was used for the nuclear phylogeny, and 10000 for the chloroplast phylogeny (lower smoothing values assume more among-branch-substitution-rate-variation than higher smoothing values). For both phylogenies, time-calibrated phylogenies were also inferred with alternative smoothing values (1, 100 and 10000 for nuclear, and 0.01, 1 and 100 for chloroplast). Experimenting with different smoothing values meant that the sensitivity of divergence time estimates to assumptions about the magnitude of among-branch-substitution-rate-variation could be determined.

Inferred divergence times within *Ipomoea* may be sensitive to the unique characteristics of different methods – especially given there are no calibrations within *Ipomoea* with which to constrain age estimates. Therefore, divergence times within *Ipomoea* were also inferred in RevBayes in two sets of analyses.

In one set of analyses, skeletal time-calibrated phylogenies for *Ipomoea* were constructed with the taxon-sampling designed such that the ages of clades of interest (set out above) were estimated. For the nuclear data, all genes that were sampled in every taxon in the skeletal phylogeny (36 taxa) and for which the aligned sequence length was at least 4,000 sites were sampled. For the chloroplast data, the entire plastomes for 37 taxa were sampled.

Divergence time estimates in the skeletal time-calibrated phylogenies may be biased by the tree prior which assumes constant diversification parameters when inferring divergence times. In the second set of analyses, a series of three taxon time-calibrated phylogenies was therefore inferred, the divergence times within which would not be affected by the assumptions of the tree prior. The taxa sampled within these three-taxon phylogenies was such that the root node of *Ipomoea* was always sampled, along with one further node from within *Ipomoea* that uniquely corresponded to a node within the skeletal phylogenies. Thus, for each node in the skeletal phylogenies, two age estimates

were generated – one derived from the skeletal phylogeny itself, and one from the three-taxon phylogeny that sampled that node. For three-taxon phylogenies constructed from nuclear data, all genes sampled for all three species in each three-taxon phylogeny were sampled. For the chloroplast data, the entire plastomes for the three species sampled in each three-taxon phylogeny were sampled.

In both the skeletal and three-taxon time-calibrated phylogenies, the MPE for the age of *Ipomoea* from Calibration Strategy 1 was used as a point calibration for the root node. In both sets of analyses, the same alignments were used as Muñoz-Rodríguez et al. (2019). For the skeletal phylogenies, the total aligned sequence length was 39,495 sites for the nuclear data, and 103,921 sites for the chloroplast data. A GTR + G + I model of sequence evolution was used, and both a strict clock and a UCLN relaxed clock. A birth-death process was used as the tree prior for the skeletal phylogenies. A Yule process was used as the tree prior for the three-taxon phylogenies. The Yule process was parameterised such that the prior probability for the age of the internal node – the single unknown divergence time – was uniformly distributed between the root age and the present. Divergence time estimates from the three-taxon phylogenies are not therefore biased by the tree prior. In all analyses, the topologies were fixed such that they were congruent with those inferred by Muñoz-Rodríguez et al. (2019). Two independent runs were performed. Sufficient mixing and convergence was assessed in Tracer v1.6.0. A 25% burn-in was used prior to calculating MAP trees, MPEs and 95% HPDs.

Inferring diversification parameters within Ipomoea

Diversification parameters were inferred in BAMM (Rabosky 2014). The 8 time-calibrated phylogenies for *Ipomoea* that were inferred in treePL (smoothing values of 0.01, 1, 100, 10000 with nuclear and chloroplast phylogenies) were used as input trees. The effect on inferred diversification parameters of different assumptions about among-branch-substitution-rate-variation (and by extension differences in divergence time estimates), and topological differences between the nuclear and chloroplast phylogenies, could therefore be determined. Following consultation with John Wood, who has carried out the majority of the taxonomic work in this project, approximate sampling proportions were specified for different clades. The set_priors.R script – downloaded with BAMM and requiring

the R package BAMMtools – was used to select the appropriate priors. Four independent runs were performed for each analysis. For the chloroplast time-calibrated phylogeny inferred with a smoothing value of 0.01, the analysis was run for 4,000,000 generations. For the chloroplast time-calibrated phylogeny inferred with a smoothing value of 1, and the nuclear time-calibrated phylogeny inferred with a smoothing value of 0.01, the analyses were run for 2,000,000 generations. In all other cases, the analyses were run for 1,000,000 generations. A 10% burnin was used. Convergence was assessed using a custom R script that required the package BAMMtools. Bayes factors were compared between different models to identify the number of rate shifts that occurred within each time-calibrated phylogeny. The credible shift set was then extracted – this represents the set of trees with different rate shift patterns that account for 95% of the posterior probability.

When did the sweet potato evolve? Did this occur in pre-human times?

To answer questions relating to the age of *I. batatas*, the MPE for the age of *Ipomoea* from Calibration Strategy 4 was used as a point calibration for the root node of time-calibrated phylogenies for *Ipomoea.* Calibration Strategy 4 does not account for the recently discovered and surprisingly old *Physalis infinemundi.* Further, several characteristics of the analysis (Magallón et al. 2015) from which Calibration Strategy 4 is derived are likely to mean that age estimates from this analysis are biased to younger ages. Age estimates derived from this calibration strategy are therefore likely to be considerably younger than those derived from Calibration Strategies 1 and 2 (and potentially 3). With respect to *I. batatas*, the primary focus of this study is to determine whether its storage root evolved in pre-human times. Biasing analyses to younger ages – by using Calibration Strategy 4 – is one of the most robust ways to do this.

The point calibration from Calibration Strategy 4 was used to calibrate the nuclear and chloroplast phylogenies for *Ipomoea* from Muñoz-Rodríguez et al. (2019). The time-calibrated phylogenies were inferred in treePL. The same optimum and alternative smoothing values were used as outlined previously.

The taxon sampling within these time-calibrated phylogenies is sufficient to estimate a divergence time between *I. batatas* and its closest wild relative *I. trifida*. Given *I. trifida* does not possess a storage root, the estimated divergence time between these two species can provide a basis for understanding when the storage root of *I. batatas* is likely to have evolved.

To scrutinise these divergence time estimates, additional time-calibrated phylogenies for *Ipomoea* sect. Batatas were inferred in RevBayes. The sensitivity of divergence time estimates to different methodological frameworks could therefore be explored.

In an initial set of time-calibrated phylogenies inferred in RevBayes, a point calibration for the root node (of *Ipomoea* sect. Batatas) was used that was equal to the age estimate for the equivalent node in the nuclear time-calibrated phylogeny (for the entire genus) inferred with the optimum smoothing value in treePL. These time-calibrated phylogenies were inferred with both nuclear and chloroplast data from Muñoz-Rodríguez et al. (2018). For the nuclear data, all genes that were sampled in every species in *Ipomoea* sect. Batatas were sampled. For the chloroplast data, the entire plastome of every species in *Ipomoea* sect. Batatas, except *I. lactifera*, was sampled. The plastome of *I. lactifera* is highly reduced, and several problems resulted from its inclusion in preliminary analyses. The same alignments were used as Muñoz-Rodríguez et al. (2018). The total aligned sequence length was 375,628 sites for the nuclear data, and 160,435 sites for the chloroplast data. For both datasets, a GTR + G + I model of sequence evolution was used, and both a strict clock and a UCLN relaxed clock. A constant rate birth-death branching process was used as the tree prior. The topologies were fixed such that they were consistent with those inferred by Muñoz-Rodríguez et al. (2018). For each time-calibrated phylogeny, two independent runs were performed. Sufficient mixing and convergence was assessed in Tracer v1.6.0. A 25% burn-in was used prior to calculating MAP trees, MPEs, and 95% HPDs.

In a further set of analyses in RevBayes, the divergence between *I. batatas* and *I. trifida* was constrained such that it occurred 15 Ka, and was therefore in a timeframe that could allow for the storage root of *I. batatas* to be a product of human domestication. This constraint meant that the substitution rates required for *I. batatas* to have evolved in human times could be inferred. Two

variants of this analysis were performed. In the first, the root node of *Ipomoea* sect. Batatas had the same point calibration as other analyses in this section. This analysis therefore assumes that the temporal framework inferred for the entire genus is broadly accurate. In the second analysis, the root node did not have a point calibration. In this analysis, there are therefore no assumptions about the accuracy of the temporal framework inferred for the entire genus. This second variant is instead entirely independent of this temporal framework.

Where the root node has a point calibration, and the divergence time between *I. batatas* and *I. trifida* is constrained to have occurred 15 Ka, major differences in substitution rates between branches may be inferred in order to satisfy these temporal constraints. As such, this analysis was performed with both a UCLN relaxed clock – where substitution rates for each branch are drawn from a single continuous distribution, and a random local clock – where discrete (and potentially very large) changes in substitution rates between different branches are explicitly modelled (Drummond and Suchard 2011). Where the root node was not fixed, a strict clock was used. This additional set of analyses was performed with only the chloroplast data. As before, a GTR + G + I model of sequence evolution was used, with a birth-death branching process as the tree prior. The topologies were fixed such that they were consistent with those inferred by Muñoz-Rodríguez et al. (2018). For each time-calibrated phylogeny, two independent runs were performed. Sufficient mixing and convergence was assessed in Tracer v1.6.0. A 25% burn-in was used prior to calculating MAP trees, MPEs, 95% HPDs.

As well as estimating the divergence time between *I. batatas* and *I. trifida*, the age of diversity within *I. batatas* was also estimated. To do this, time-calibrated phylogenies with multiple accessions of *I. batatas* (71 accessions) and *I. trifida* (22 accessions) were inferred. These time-calibrated phylogenies were inferred in treePL and RevBayes.

For analyses in treePL, nuclear and chloroplast phylogenies from Muñoz-Rodríguez et al. (2018), that contain multiple accessions of *I. batatas* and *I. trifida*, were used as input trees. A point calibration was used for the root node that was equal to the age estimate for the equivalent node in the nuclear time-calibrated phylogeny for the entire genus inferred with the optimum smoothing value in treePL. Following extensive cross-validation, a smoothing value of 0.01 was used for the nuclear and

chloroplast phylogenies. Time-calibrated phylogenies were also inferred with alternative smoothing values (1, 100, 10000) to determine the sensitivity of divergence time estimates to different assumptions about the magnitude of among-branch-substitution-rate-variation.

In RevBayes, two sets of analyses were performed. First, time-calibrated phylogenies were inferred for all sampled accessions with either the nuclear or chloroplast data. For the nuclear data, the 40 genes that were sampled in every accession and had the longest aligned sequence length were sampled. For the chloroplast data, entire plastomes were sampled. Second, a series of three-taxon time calibrated phylogenies were inferred with either the nuclear or chloroplast data. The motivation for this was the same as set out for the three-taxon time-calibrated phylogenies inferred across the whole of *Ipomoea*. The taxon sampling was such that in each three-taxon phylogeny, the root node of the phylogeny containing all sampled accessions was always sampled, along with one further node. For the nuclear data, all genes were sampled that were sampled for all three species in each three-taxon phylogeny. For the chloroplast data, the entire plastome of the three species sampled in each three-taxon phylogeny were sampled. The same alignments as Muñoz-Rodríguez et al. (2018) were used.

In all these analyses, a GTR + G + I model of sequence evolution was used with either a strict clock or a UCLN relaxed clock. A point calibration for the root node was used that was equal to the age estimate for the equivalent node in the nuclear time-calibrated phylogeny for the entire genus inferred with the optimum smoothing value in treePL. For the time-calibrated phylogenies containing all sampled accessions, a birth-death branching process was used as the tree prior. For the three-taxon time-calibrated phylogenies, a Yule process was used as the tree prior. This was parameterised such that the prior probability for the age of the internal node – the single unknown divergence time – was uniformly distributed between the root age and the present. In all analyses, the topologies were fixed such that they were consistent with those inferred by Muñoz-Rodríguez et al. (2018). For each time-calibrated phylogeny, two independent runs were performed. Sufficient mixing and convergence was assessed in Tracer v1.6.0. A 25% burn-in was used prior to calculating MAP trees, MPEs and 95% HPDs.

All custom R, Python, and RevBayes scripts developed for this study are available in Electronic Appendix 3.

RESULTS

Time-calibrated phylogenies for Solanaceae and Convolvulaceae, and an age estimate for Ipomoea

Here, I summarise the results when a UCLN relaxed clock is used with different calibration strategies for Convolvulaceae and Solanaceae. A comparison with results when using a strict clock follows.

Calibration Strategy 1 resulted in the oldest divergence time estimates throughout Convolvulaceae and Solanaceae (Fig. 4.1a). The 95% HPD for the age of the divergence between Convolvulaceae and Solanaceae was 97-130 Ma, the 95% HPD for the age of the Convolvulaceae crown node was 94-129 Ma, the 95% HPD for the age of the Solanaceae crown node was 67-99 Ma, and the 95% HPD for the age of the *Ipomoea* crown node was 18-43 Ma. There was a 20.3-fold difference between the maximum inferred substitution rate (MPE) of $2.9*10^{-3}$ substitutions Myr^{-1} (on the Solanaceae stem branch), and the minimum inferred substitution rate of $1.4*10^{-4}$ substitutions Myr^{-1} (Fig. 4.1a and Appendix 4). However, apart from the Solanaceae stem branch, inferred rates are within a range between $1.4*10^{-4}$ and $1.4*10^{-3}$ substitutions Myr^{-1}, a 10-fold difference.

Calibration Strategy 2 resulted in younger divergence time estimates (Fig. 4.1b). The 95% HPD for the age of the divergence between Convolvulaceae and Solanaceae was 89-129 Ma, the 95% HPD for the age of the Convolvulaceae crown node was 86-126 Ma, the 95% HPD for the age of the Solanaceae crown node was 64-94 Ma, and the 95% HPD for the age of the *Ipomoea* crown node was 17-41 Ma. There was a 21.5-fold difference between the maximum inferred substitution rate of $3.2*10^{-3}$ substitutions Myr^{-1} (on the Solanaceae stem branch), and the minimum inferred substitution rate of $1.5*10^{-4}$ substitutions Myr^{-1} (Fig. 4.1b and Appendix 4). However, apart from the Solanaceae stem branch, inferred rates fall within a range between $1.5*10^{-4}$ and $1.5*10^{-3}$ substitutions Myr^{-1}, a 10-fold difference.

Calibration Strategy 3 resulted in the youngest divergence time estimates (Fig. 4.1). The 95% HPD for the age of the divergence between Convolvulaceae and Solanaceae was 57-58 Ma, the 95% HPD for the age of the Convolvulaceae crown node was 46-56 Ma, the 95% HPD for the age of the Solanaceae crown node was 56-58 Ma, and the 95% HPD for the age of the *Ipomoea* crown node was 10-21 Ma. There was a 59-fold difference between the maximum inferred substitution rate of $1.3*10^{-2}$ substitutions Myr^{-1} (on the Convolvulaceae stem branch), and the minimum inferred substitution rate of $2.1*10^{-4}$ substitutions Myr^{-1} (Fig. 4.1c and Appendix 4). However, apart from the Convolvulaceae stem branch, inferred rates fall within a range between $1.5*10^{-4}$ and $3.5*10^{-3}$ substitutions Myr^{-1}, a 16-fold difference.

Calibration Strategy 4 also resulted in younger divergence time estimates than Calibration Strategy 1 and 2 (Fig. 4.1). The 95% HPD for the age of the divergence between Convolvulaceae and Solanaceae was 41-83 Ma, the 95% HPD for the age of the Convolvulaceae crown node was 40-81 Ma, the 95% HPD for the age of the Solanaceae crown node was 26-59 Ma, and the 95% HPD for the age of the *Ipomoea* crown node was 8-25 Ma. There was a 17.9-fold difference between the maximum inferred substitution rate of $4.8*10^{-3}$ substitutions Myr^{-1} (on the Solanaceae stem branch), and the minimum inferred substitution rate of $2.7*10^{-4}$ substitutions Myr^{-1} (Fig. 4.1d and Appendix 4). However, apart from the Solanaceae stem branch, inferred rates fall within a range between $1.5*10^{-4}$ and $2.5*10^{-3}$ substitutions Myr^{-1}, a 9.2-fold difference.

With a strict clock, Calibration Strategies 1-3 resulted in divergence time estimates that were very similar to each other, and to divergence time estimates with Calibration Strategy 1 and a UCLN relaxed clock (Appendix 5; Fig. 4.1a). Inferred substitution rates were also very similar for Calibration Strategies 1-3 with a strict clock (Appendix 5). With a strict clock, Calibration Strategy 4 resulted in markedly younger divergence time estimates than Calibration Strategies 1-3 with a strict clock and calibration Strategy 4 with a UCLN relaxed clock (Appendix 5, Fig. 4.1d). For all four calibration strategies, 95% HPDs were far narrower with a strict clock (Fig. 4.1, Appendix 5).

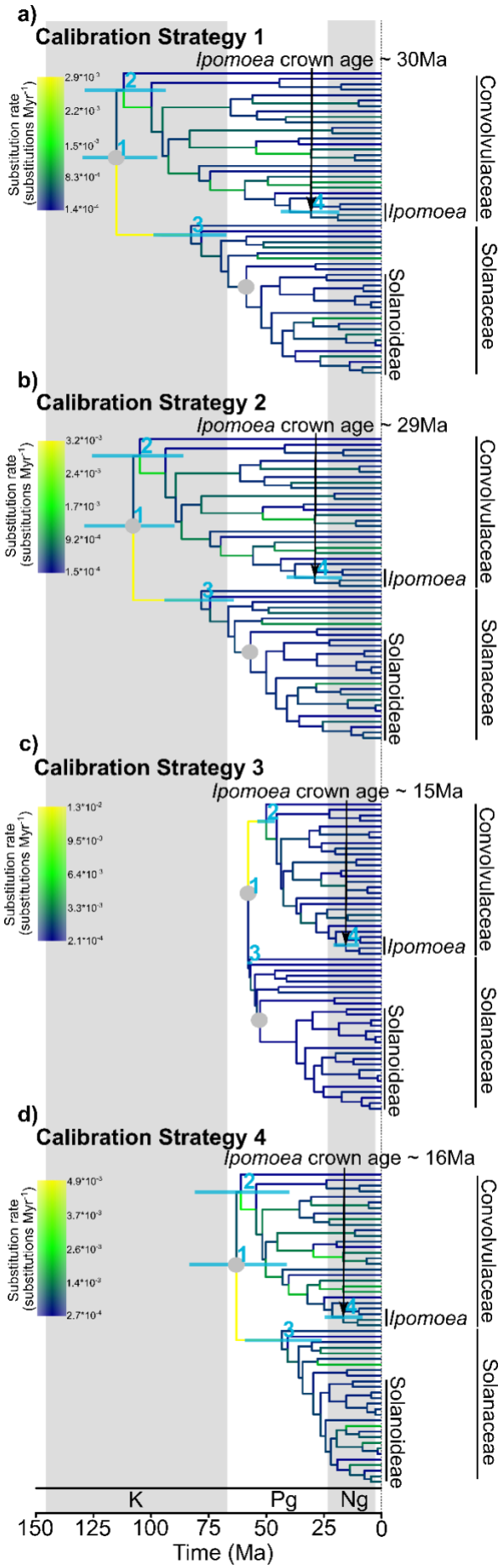

Figure 4.1. Time-calibrated phylogenies for Convolvulaceae and Solanaceae inferred with different calibration strategies and a UCLN relaxed clock. Branch colours refer to the inferred substitution rate (MPE). Grey circles refer to calibrated nodes, either the Solanaceae stem node or the Solanoideae stem node. Blue bars with numbers refer to the 95% HPD for key nodes; 1 refers to the divergence between Convolvulaceae and Solanaceae; 2 refers to the Convolvulaceae crown node; 3 refers to the Solanaceae crown node; 4 refers to the *Ipomoea* crown node. a) Calibration Strategy 1 – the Solanaceae stem node has a maximum age constraint of 130 Ma, and the Solanoideae stem node has a minimum age constraint of 52 Ma. b) Calibration Strategy 2 – the Solanaceae stem node is calibrated with a lognormal distribution with offset = 52, m = 10.4, v =10.4. The Solanoideae stem node has a minimum age constraint of 52 Ma. c) Calibration Strategy 3 – the Solanaceae stem node is calibrated with a lognormal distribution with offset = 52, m = 5.2, v =2.1. The Solanoideae stem node has a minimum age constraint of 52 Ma. d) The Solanaceae stem node is calibrated with a secondary calibration from Magallón et al. 2015. This is implemented as a normal distribution with $\mu = 67.34, \sigma = 9.980$.

Divergence time estimates within Ipomoea

Divergence time estimates within Ipomoea inferred in treePL with the optimum smoothing value

When the optimum smoothing value of 0.01 was used to infer a nuclear time-calibrated phylogeny in treePL, the inferred crown ages for the New World and Old World Clades were approximately 29 Ma and 28 Ma respectively (Fig. 4.2a, Appendix 6). The major clades A-E, and *Astripomoea* and close relatives, are inferred to have originated 16-23 Ma, throughout the Early Miocene (Fig. 4.2a, Appendix 6). The Parana Clade and Coriaceous Sepal Clade, and *Ipomoea* sect. Batatas are inferred to have originated 7.5-12.5 Ma, during the mid Miocene. In this time-calibrated phylogeny, there was a 3.279-fold difference between the minimum inferred substitution rate of $6.21*10^{-4}$ substitutions Myr^{-1}, and the maximum inferred substitution rate of $2.036*10^{-3}$ substitutions Myr^{-1} (Fig. 4.2a, Appendix 8).

When the optimum smoothing value of 10000 was used to infer a chloroplast time-calibrated phylogeny in treePL, the inferred crown ages for the New World and Old World Clades were approximately 25.6 Ma and 21 Ma respectively (Fig. 4.2b, Appendix 6). Clades A, B, D, E, and *Astripomoea* and close relatives, are inferred to have originated 13.5-19.1 Ma, throughout the Early and Middle Miocene. Clade C is inferred to have originated during the mid Miocene, 8.9 Ma (Fig. 4.2b, Appendix 6). In this time-calibrated phylogeny, the same substitution rate of $1.81*10^{-4}$ substitutions Myr^{-1} was inferred for every branch (Fig. 4.2b).

These results from the nuclear and chloroplast time-calibrated phylogenies indicate that much of the morphological and species diversity within *Ipomoea*, briefly summarised in Figure 4.2c-j, originated before the end of the Miocene. However, the limited taxon sampling means that it is difficult to construct detailed or precise hypotheses about when and how frequently specific morphological traits evolved.

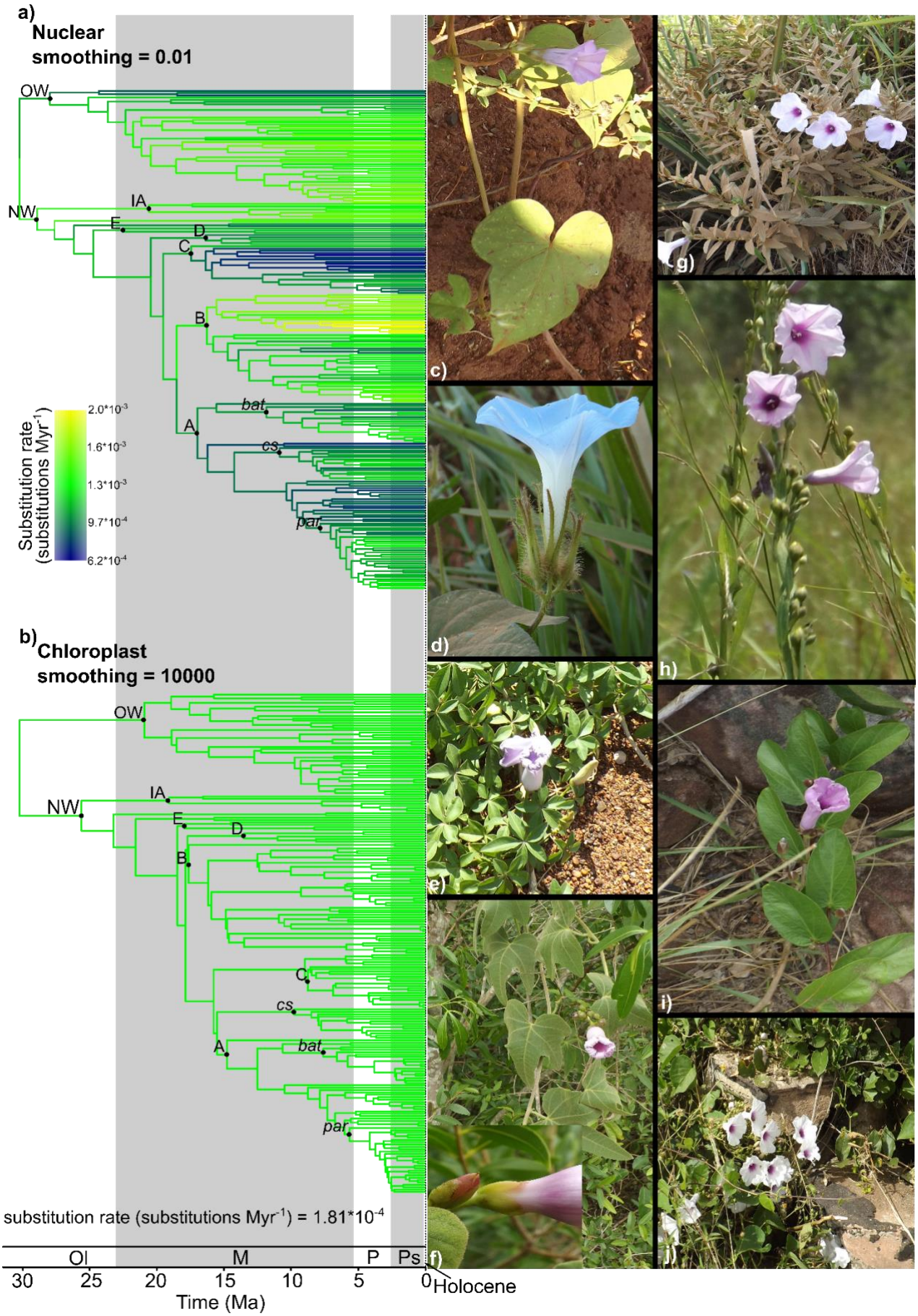
a)
Nuclear
smoothing = 0.01
OW
NW
IA
E
D
C
B
A
bat
cs
par
Substitution rate
(substitutions Myr-1)
2.0*10-3
1.6*10-3
1.3*10-3
9.7*10-4
6.2*10-4
b)
Chloroplast
smoothing = 10000
OW
NW
IA
E
D
B
C
cs
A
bat
par
substitution rate (substitutions Myr-1) = 1.81*10-4
Ol
M
P
Ps
30
25
20
15
10
5
0
Time (Ma)
Holocene
c)
g)
d)
h)
e)
i)
f)
j)

Figure 4.2. (previous page) Time-calibrated phylogenies for *Ipomoea* inferred in treePL with the optimum smoothing value, and images of species from different clades. A point calibration of 30.3 Ma is used at the root node in both a) and b). This is MPE for the age of *Ipomoea* from Calibration Strategy 1. Where rates differ between branches, branch colours refer to the inferred substitution rate. Labelled nodes refer to nodes defining major clades. A – E refer to clades A – E from Muñoz-Rodríguez et al. (2019) (see methods). *par* refers to the Parana clade, *cs* refers to the Coriaceous Sepal clade, *bat* refers to *Ipomoea* sect. Batatas, and IA refers to *Astripomoea* and close relatives. a) shows a nuclear time-calibrated phylogeny, and b) shows a chloroplast time-calibrated phylogeny. c) *I. cordatotriloba*, a trailing and twining species in *Ipomoea* sect. Batatas. d) *I. nil*, a twining species in clade B. e) *I. cairica*, a trailing and twining species in clade E. f) *I. bonariensis*, a trailing and twining species in the Coriaceous Sepal Clade. g) *I. paraguariensis*, an erect species in the Parana clade. h) *I. paludosa*, an erect species in the Parana clade. i) *I. psammophila*, a trailing species in the Parana clade. j) *I. megapotamica*, a trailing and climbing species in the Parana clade.

Alternative divergence time estimates for the Ipomoea nuclear phylogeny

Divergence times inferred in treePL with alternative smoothing values were similar to divergence times inferred with the optimum value, although in some cases they were somewhat younger (Appendix 6 and 7, Table 4.1). With higher smoothing values, the magnitude of among-branch-substitution-rate-variation decreased (Appendix 7 and 8).

For analyses in RevBayes, in the skeletal time-calibrated phylogeny inferred with a strict clock, MPEs for the ages of major clades were similar to age estimates inferred in treePL (Appendix 6 and 7, Table 4.1). 95% HPD intervals were on average 10.9% of the MPE for the age each clade (Appendix 6, Table 4.1), and the inferred substitution rate was similar to those inferred in treePL (Appendix 7). In the three-taxon time-calibrated phylogenies inferred with a strict clock, MPEs for the ages of major clades were somewhat older than in the skeletal time-calibrated phylogeny and 95% HPDs were on average 4.7% of the MPE for the age of each clade (Appendix 6, Table 4.1).

In the skeletal time-calibrated phylogenies inferred with a relaxed clock, MPEs for the ages of major clades were similar to age estimates inferred in treePL and to the skeletal time-calibrated phylogeny inferred with a strict clock. However, there was more variation in the relationship between these age estimates and other estimates, compared to those previously discussed, with inferred ages being either older or younger depending on the clade (Appendix 6 and 7, Table 4.1). 95% HPD intervals were markedly wider than in RevBayes analyses with a strict clock, on average they were 60% of the MPE for the age of each clade (Appendix 6, Table 4.1). The substitution rate and

magnitude of among-branch-substitution-rate-variation was similar to that inferred in treePL with a lower smoothing value (Appendix 7 and 8). In three-taxon time-calibrated phylogenies inferred with a relaxed clock, MPEs tended to be slightly older than MPEs inferred in either of the skeletal time-calibrated phylogenies. However, for the two oldest clades in the phylogeny, NW and OW, MPEs were markedly younger than other age estimates inferred in treePL or RevBayes (Appendix 6, Table 4.1). 95% HPD intervals were considerably wider compared to the skeletal time-calibrated phylogeny inferred with a relaxed clock. On average they were 114.1% of the MPE for the age of each clade (Appendix 6, Table 4.1).

Alternative divergence time estimates for the Ipomoea chloroplast phylogeny

Divergence times inferred in treePL with a smoothing value of 100 were very similar to divergence times inferred with a smoothing value of 10000 (Appendix 6 and 7, Table 4.1). With a smoothing value of 0.01 or 1, inferred divergence times were significantly older. This was especially the case for clades A-E, the majority of which were inferred to have originated during the earliest Miocene or late Oligocene (Appendix 6 and 7, Table 4.1). With lower smoothing values, the magnitude of among-branch-substitution-rate-variation increased (Appendix 7 and 8). With the lowest smoothing value of 0.01, there was a 193-fold difference between the maximum inferred substitution rate of $3.86*10^{-4}$ substitutions Myr^{-1}, and the minimum inferred substitution rate of $2*10^{-6}$ substitutions Myr^{-1}.

For analyses in RevBayes, in skeletal time-calibrated phylogenies inferred with a strict clock, MPEs for the ages of major clades were most similar to age estimates inferred in treePL with a smoothing value of 10000 (Appendix 6, Appendix 7, Table 4.1). 95% HPD intervals were approximately 16% of the MPE for the age of each clade (Appendix 6, Table 4.1). As with divergence times, the inferred substitution rate was also similar to when a smoothing value of 10000 was used in treePL (Appendix 7). In three-taxon time calibrated phylogenies inferred with a strict clock, MPEs for the ages of major clades were similar to those inferred in the skeletal time-calibrated phylogeny (Appendix 6, Table 4.1). 95% HPDs were slightly wider than in the skeletal time-calibrated phylogeny, approximately 19% of the MPE for the age of each clade (Appendix 6, Table 4.1).

In skeletal time-calibrated phylogenies inferred with a relaxed clock, MPEs for the ages of major clades tended to be slightly older than in either of the two types of time-calibrated phylogeny inferred with a strict clock (Appendix 6 and 7, Table 4.1). They were therefore also slightly closer to age estimates inferred in treePL with smoothing values of 0.01 and 1. 95% HPDs were also markedly wider than RevBayes analyses with a strict clock, approximately 72% of the MPE for the age of each clade (Appendix 6, Table 4.1). The substitution rate, and magnitude of among-branch-substitution-rate-variation was similar to the chloroplast time-calibrated phylogeny inferred in treePL with a smoothing value of 0.01 (Appendix 7 and 8). In three-taxon time-calibrated phylogenies inferred with a relaxed clock, MPEs for the ages of major clades tended to be slightly older than MPEs inferred according to alternative methods in RevBayes (Appendix 6, Table 4.1). 95% HPD intervals were also markedly wider, approximately 132.4% of the MPE for the age of each clade (Appendix 6, Table 4.1).

Table 4.1. Results summary of time-calibrated phylogenies for *Ipomoea*

	Inference Method	**% Change in age estimate relative treePL with smoothing = 0.01**	**Mean 95% HPD width as % of MPE**
Nuclear	**Smoothing = 0.01**	--	--
	Smoothing = 1	-10.1	--
	Smoothing = 100	-10.3	--
	Smoothing = 10000	-10.3	--
	Skeletal strict clock	-15.3	10.9
	Three-taxon strict clock	7.2	4.7
	Skeletal UCLN relaxed clock	-12.3	60.0
	Three-taxon UCLN relaxed clock	7.0	114.1
Chloroplast	**Smoothing = 0.01**	--	--
	Smoothing = 1	4.1	--
	Smoothing = 100	-27.3	--
	Smoothing = 10000	-27.6	--
	Skeletal strict clock	-26.8	15.9
	Three-taxon strict clock	-23.6	18.9
	Skeletal UCLN relaxed clock	-22.3	71.5
	Three-taxon UCLN relaxed clock	-18.7	132.4

Diversification parameter estimates within Ipomoea

Nuclear Phylogeny

Inferred diversification parameters were similar, regardless of the smoothing value that was used to infer the time-calibrated phylogeny (Fig. 4.3a and Appendix 9). With the time-calibrated phylogeny inferred with the optimum smoothing value of 0.01, a rate shift pattern with a single diversification rate shift near the origin of a clade that contains the Parana Clade and Coriaceous Sepal Clade has by far the highest posterior probability (p = 0.9) (Fig. 4.3a). This diversification rate shift leads to a 5.9-fold increase in the net diversification rate to 0.802 species Myr^{-1}, compared to the background rate of 0.136 species Myr^{-1}. This inferred net diversification rate increase is caused by an increase in the inferred speciation rate from 0.147 to 0.820 species Myr^{-1}.

Chloroplast Phylogeny

In the time-calibrated phylogeny inferred with the optimum smoothing value of 10000, a rate shift pattern with a single net diversification rate shift near the origin of the Parana Clade has the highest posterior probability (p = 0.79) (Fig. 4.3b). Two alternative rate shift patterns, in which there are net diversification rate shifts on immediately ancestral branches, have significantly lower posterior probabilities (Fig. 4.3b). The most probable diversification rate shift pattern leads to a 3.2-fold increase in net the diversification rate to 0.641 species Myr^{-1} , compared to the background rate of 0.199 species Myr^{-1}. This inferred net diversification rate increase is caused by an increase in the inferred speciation rate from 0.212, to 0.768 species Myr^{-1}. With the time-calibrated phylogeny inferred with a smoothing value of 100, inferred diversification parameters were very similar to when a smoothing value of 10000 was used (Appendix 9). With the time-calibrated phylogenies inferred with smoothing values of 1 and 0.01, no discrete rate shifts were inferred (Appendix 9).

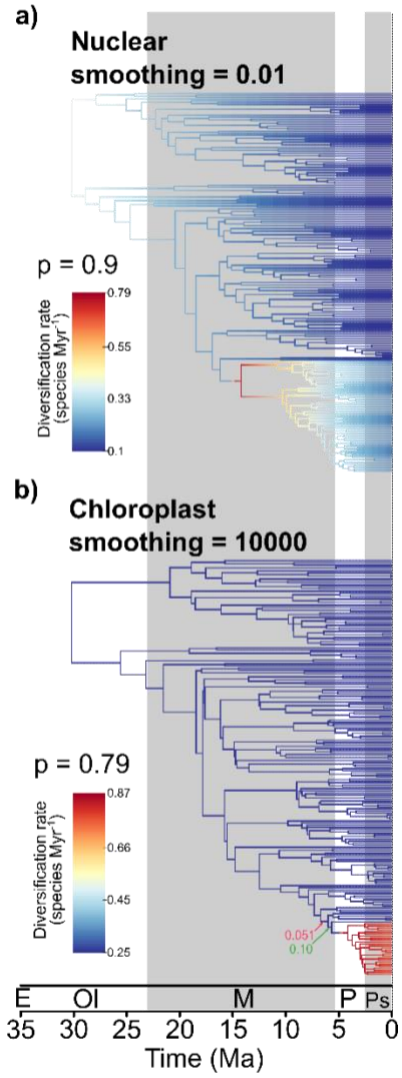

Figure 4.3. Net diversification rates for *Ipomoea* inferred in BAMM. Time-calibrated phylogenies inferred in treePL with the optimum smoothing values are used as input trees. The net diversification rate shift pattern with the highest posterior probability for each time-calibrated phylogeny is shown as the coloured tree. If present, alternative rate shift patterns that have a posterior probability of greater than 0.05 are indicated with coloured arrows, with the position of the arrow indicating the position of the rate shift. Arrows with the same colour indicate a single set of rate shifts that have a specific posterior probability (indicated next to one arrow for each colour). In a) diversification parameters are estimated with the nuclear time-calibrated phylogeny. In b) diversification parameters are estimated with the chloroplast time-calibrated phylogeny.

When did the sweet potato evolve?

In time-calibrated phylogenies for *Ipomoea* sect. Batatas inferred in RevBayes with a UCLN relaxed clock, the divergence between *I. batatas* and *I. trifida* was inferred to have occurred at least 1.9 Ma in the nuclear time-calibrated phylogeny (Fig. 4.4a), and at least 0.89 Ma in the chloroplast time-calibrated phylogeny (Fig. 4.4b). In nuclear and chloroplast time-calibrated phylogenies inferred with alternative methodologies, the inferred divergence time between *I. batatas* and *I. trifida* was always inferred to have occurred at least 0.8 Ma (Appendix 10). These alternative methodologies make different assumptions about the magnitude and nature of among-branch-substitution-rate-variation. As such, each analysis makes different inferences about substitution rates on individual

branches (Fig. 4.4, Appendix 10). Regardless of this, in all cases the divergence between *I. batatas* and *I. trifida* is inferred to have occurred in pre-human times.

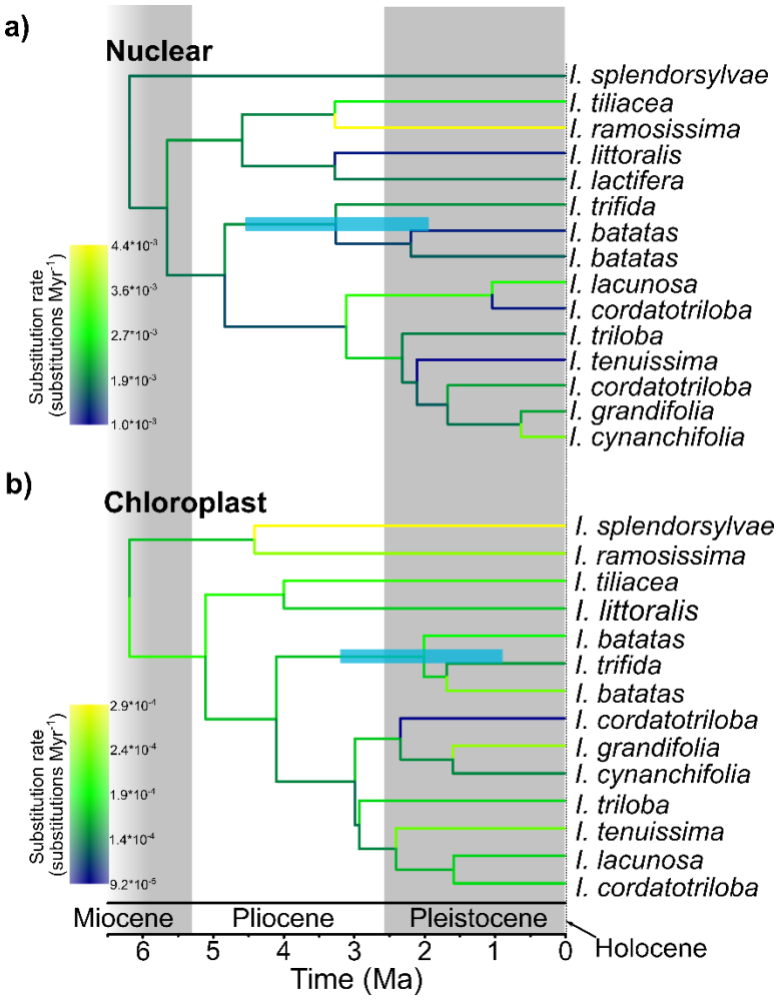

Figure 4.4. Time-calibrated phylogenies for *Ipomoea* sect. *Batatas* inferred in RevBayes with a UCLN relaxed clock. Branch colours refer to the inferred substitution rate. The blue bar refers to the 95% HPD for the divergence time between *I. batatas* and *I. trifida*. a) shows the nuclear time-calibrated phylogeny, b) shows the chloroplast time-calibrated phylogeny.

Two further analyses were performed with the chloroplast dataset. Where the divergence between *I. batatas* and *I. trifida* was constrained to have occurred no more than 15 Ka, but the point calibration of 6.2 Ma was retained at the root node – a greater than 200-fold increase in substitution rates for branches leading to *I. batatas* and *I. trifida* was inferred. As a result, substitution rates in this part of the time-calibrated phylogeny exceed $2.2*10^{-2}$ substitutions Myr^{-1} (Fig. 4.5a, Appendix 11). This value far exceeds any previously inferred substitution or mutation rate for chloroplast DNA. This result was consistent regardless of whether a UCLN relaxed clock (Fig. 4.5a) or a random local clock (Appendix 11) was used to infer substitution rates.

Where the divergence between *I. batatas* and *I. trifida* was constrained to have occurred 15 Ka, but the point calibration at the root node was removed, a substitution rate of $2.1*10^{-2}$ substitutions Myr^{-1} was inferred for *Ipomoea* sect. Batatas (Fig. 4.5b). As before, this rate is far higher than any previously inferred substitution or mutation rate for chloroplast DNA. Taken together, these two additional analyses demonstrate that in order to infer a divergence time between *I. batatas* and *I. trifida* that occurred in human times, a substitution rate would be required that far exceeds any known substitution or mutation rate for chloroplast DNA.

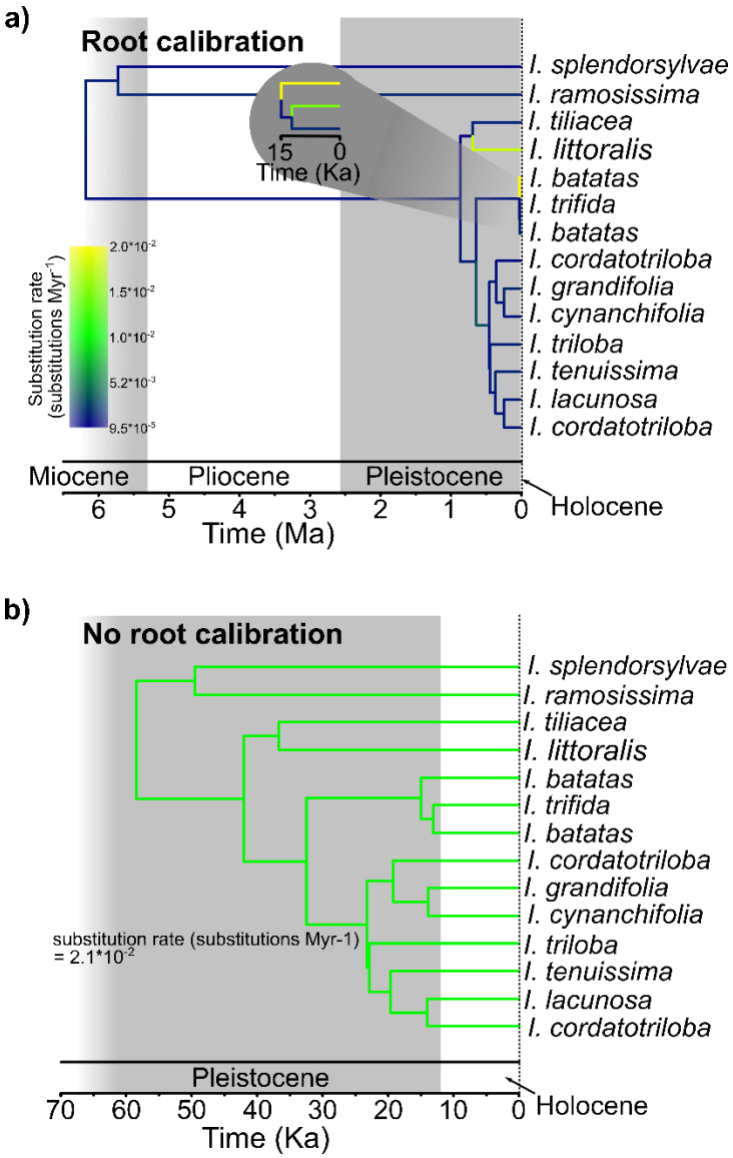

Figure 4.5. Alternative chloroplast time-calibrated phylogenies for *Ipomoea* sect Batatas inferred in RevBayes. Where rates differ between branches, branch colours refer to the inferred substitution rate. In a) the divergence between *I. batatas* and *I. trifida* is calibrated to 15 Kyr and the root node is calibrated to 6.18 Ma – the same root calibration as was used in all the previous time-calibrated phylogenies for *Ipomoea* sect. Batatas inferred in RevBayes. A UCLN relaxed clock is used to infer substitution rates. In b) the divergence between *I. batatas* and *I. trifida* is constrained to have occurred 15 Ka and there is no calibration at the root node. A strict clock is used to infer substitution rates.

When did diversity within the sweet potato evolve?

With the time-calibrated phylogenies that contained all sampled accessions of *I. batatas* and *I. trifida*, the divergences between most of the lineages within *I. batatas* occurred in pre-human times (Fig. 4.6, Table 4.2). This represents strong evidence that most of the diversity within *Ipomoea batatas* evolved in pre-human times. This result was consistent with both the nuclear (Fig. 4.6a, Table 4.2) and chloroplast datasets (Fig. 4.6b, Table 4.2), and for analyses inferred in treePL with both the optimum and alternative smoothing values, or in RevBayes with a strict clock or UCLN relaxed clock (Appendix 12, Table 4.2). These alternative methodologies make different assumptions about the magnitude and nature of among-branch-substitution-rate-variation. As such, each analysis makes different inferences about substitution rates on individual branches (Fig 4.6, Appendix 12 and 13, Table 4.2). Regardless of this, in nearly all cases the divergences between different accessions of *I. batatas* and *I. trifida* are inferred to have occurred in pre-human times (Fig. 4.6, Appendix 12, Table 4.2).

In RevBayes, divergence times were also inferred in three-taxon time-calibrated phylogenies (Table 4.2). The divergence time estimates in these three-taxon time-calibrated phylogenies differed somewhat from those in which all taxa were sampled, and HPDs were considerably wider (Table 4.2, Appendix 14). However, as before, the divergences between different accessions of *I. batatas* and *I. trifida* predominantly occurred in pre-human times (Table 4.2; Appendix 14).

Table 4.2. Results summary of time-calibrated phylogenies inferred in RevBayes for estimating the age of diversity within *I. batatas* and *I. trifida*

		% change in MPE in three-taxon time-calibrated phylogeny compared to time-calibrated-phylogeny with all specimens	Mean 95% HPD width (as % of MPE) – all specimens	Mean 95% HPD width (as % of MPE) – three taxon	% of nodes with 95% HPD that includes human era (< 15 Ka) – all specimens	% of nodes with 95% HPD that includes human era (< 15 Ka) – three taxon
Nuclear	**Strict clock**	29.0	13.5	12.7	0	1.5
	UCLN relaxed clock	-16.0	41.6	124.3	0	0
Chloroplast	**Strict clock**	33.9	65.4	94.0	0	5.9
	UCLN relaxed clock	33.5	111.1	183.7	1.5	2.9

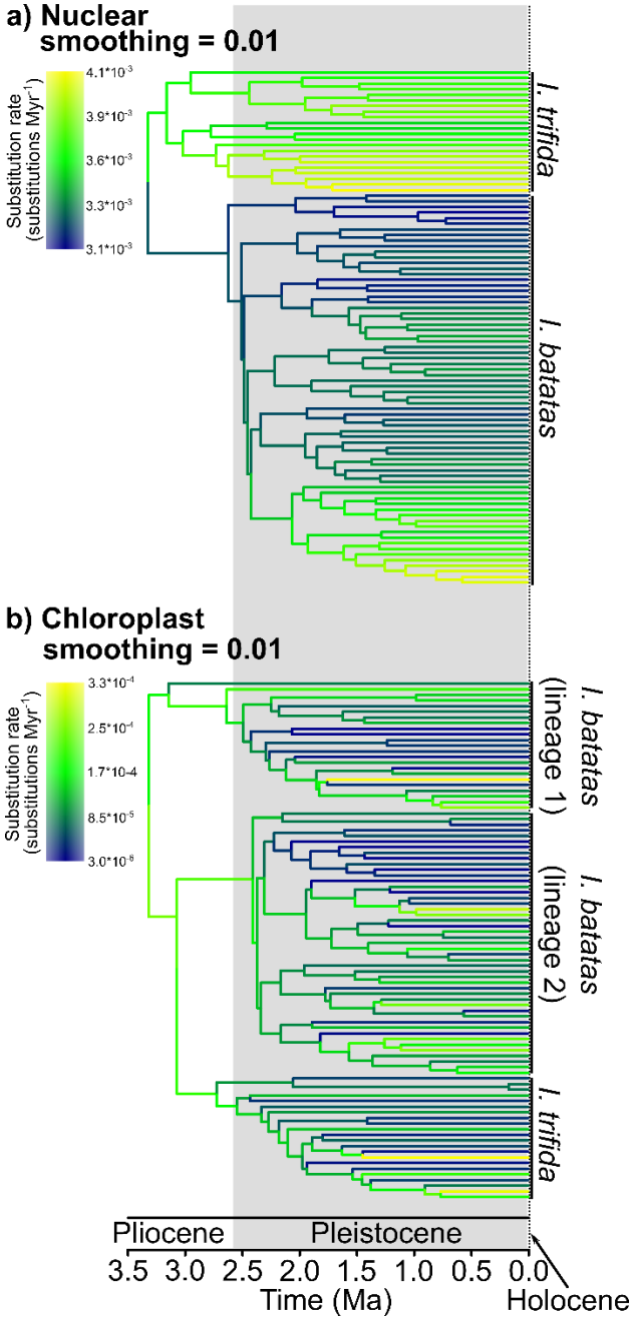

Figure 4.6. Time-calibrated phylogenies for *I. batatas* and *I. trifida* inferred in treePL with the optimum smoothing value. All sampled accessions of the two species are included. Branch colours refer to the inferred substitution rate. a) shows the nuclear time-calibrated phylogeny, b) shows the chloroplast time-calibrated phylogeny.

DISCUSSION

Calibration uncertainty provides an uncertain basis for inferring the temporal dynamics of evolution in Ipomoea

Different combinations and interpretations of fossil calibrations provide an uncertain and conflicting framework for calibrating the phylogeny of Convolvulaceae and Solanaceae. As such, different calibration strategies led to major differences in inferred divergence times within the two

families, including for the crown node of *Ipomoea* (Fig. 4.1). With respect to fossils within Solanaceae, this study focussed on the recently discovered *Physalis infinemundi*, and the implications of different assumptions about the maximum possible age of Solanaceae. It was shown that making stronger assumptions that the Solanaceae stem node is significantly younger than 130 Ma – a commonly used hard maximum for the origin of eudicots – led to significantly younger divergence time estimates. This was the case even when *Physalis infinemundi* was retained as a minimum constraint of 52 Ma for the Solanoideae stem node. In Calibration Strategy 3, which makes the strongest assumption that the Solanaceae stem node is significantly younger than 130 Ma, age estimates were similar to those obtained in previous studies where significantly younger fossil calibrations were used (Särkinen et al. 2013; Magallón et al. 2015; Mitchell et al. 2016). As well as evaluating fossil calibrations within Solanaceae, a time-calibrated phylogeny was also constructed with a secondary calibration from Magallón et al. (2015). This led to younger divergence time estimates that were similar to Calibration Strategy 3.

It is impossible to make a definitive judgement about which calibration strategy is the most robust. Assuming its phylogenetic placement is correct, *Physalis infinemundi* does provide a robust minimum constraint for the Solanoideae stem node. However, it would be misleading to suggest we know a most probable age for the Solanoideae stem node in relation to the age of this fossil. Alternatively, it may be reasonable to suggest that the Solanaceae stem node must be younger than 130 Ma – but the likely extent to which it is younger is also difficult to quantify. Further, the maximum constraint of 130 Ma is derived from the observation that there is an absence of eudicot fossils in sediments older than this age, but a sudden appearance in younger sediments (Doyle et al. 1977). This maximum constraint is therefore based on the absence of evidence – a problematic basis upon which to construct assumptions.

The implications of different calibration strategies are also influenced by assumptions about among-branch-substitution-rate-variation. Note, for example, how the same calibration strategy can lead to significantly different divergence time estimates, depending on whether a UCLN relaxed clock or a strict clock is used (Fig. 4.1 and Appendix 5). Further, for the same molecular clock model,

different calibration strategies lead to marked differences in inferred substitution rates (Fig. 4.1, Appendix 4 and 5). Taken together, this highlights the tight interaction between fossil calibrations, assumptions about among-branch-substitution-rate-variation, inferred substitutions rates, and inferred divergence times.

Despite the uncertainty presented here, I suggest that presenting results in this way provides the most robust basis for interpreting divergence time estimates, and accounting for factors that may affect inferences in the future. I do not, by contrast, present a single time-calibrated phylogeny, because no single time-calibrated phylogeny fully accounts for all the uncertainty that is represented here. Despite this complexity, with careful interpretation these results can provide a basis for discussing the temporal dynamics of evolution in *Ipomoea*, and answering specific questions about the origin of *I. batatas*.

Uncertainty about substitution rate variation within Ipomoea leads to further uncertainty in divergence time estimates

No fossil calibrations were implemented at internal nodes within *Ipomoea*. Therefore, whilst uncertainty about the absolute age of the entire of *Ipomoea* can result from calibration uncertainty (summarised in Figure 4.1), uncertainty about the relative ages of different clades within *Ipomoea* cannot result from calibration uncertainty. Instead, it results predominantly from assumptions about substitution rates, and in some cases additional assumptions about relative clade ages (expressed by the birth-death branching process in Bayesian analyses). In analyses of divergence times within *Ipomoea* presented here, I therefore experimented with different assumptions about substitution rates, and in Bayesian analyses, with the assumptions of the birth-death branching process.

In nuclear time-calibrated phylogenies inferred in treePL, different smoothing values have a limited effect on inferred divergence times (Fig. 4.2a, Appendix 6 and 7). Regardless of the smoothing value, the majority of major clades are inferred to have originated in the Late Oligocene and throughout the early Miocene, and the Parana Clade, Coriaceous Sepal Clade, and *Ipomoea* sect. Batatas are inferred to have originated in the Late Miocene. This implies that inferred divergence

times within *Ipomoea* are fairly robust to assumptions about among-branch-substitution-rate-variation. This result is likely to be explained by the fact that regardless of the smoothing value, the magnitude of inferred among-branch-substitution-rate-variation was always fairly low (Fig. 4.2a, Appendix 8). Nonetheless, any conclusion about the robustness of these divergence time estimates assumes that the model of rate variation used in treePL – whereby substitution rates are correlated between ancestral and descendant branches (Sanderson 1997; Smith and O'Meara 2012) – is valid.

By contrast, in chloroplast time-calibrated phylogenies inferred in treePL, different smoothing values had a more profound effect on inferred divergence times. Specifically, lower smoothing values (0.01 and 1) caused basal branches to be considerably shorter – and by consequence have considerably higher substitution rates – compared to higher smoothing values. This in turn meant that with lower smoothing values, inferred divergence times were considerably older such that the majority of major clades originated in the Late Oligocene or very early Miocene (Fig. 4.2b, Appendix 6 and 7, Table 4.1). Clearly, this result highlights that there is considerable uncertainty in divergence time estimates within *Ipomoea*. However, when experimenting with different smoothing values, extensive cross-validation analyses were performed to determine which smoothing values are most appropriate. For the chloroplast data, these suggested that higher smoothing values are more appropriate. Further, the divergence times inferred in the chloroplast time-calibrated phylogeny with higher smoothing values are more congruent with the nuclear time-calibrated phylogeny, with major clades originating throughout the Early Miocene, rather than exclusively in the very early Miocene. I therefore tentatively suggest that divergence times inferred for the chloroplast dataset with higher smoothing values are more accurate (Fig. 4.2b).

For analyses performed with nuclear data in RevBayes, the skeletal and three-taxon time-calibrated phylogenies were broadly congruent with those inferred in treePL (Appendix 6 and 7, Table 4.1). This provides further evidence that divergence time estimates with this dataset are relatively insensitive to assumptions about among-branch-substitution-rate-variation. However, 95% HPDs are considerably wider when a UCLN relaxed clock is used compared to a strict clock (Appendix 6, Table 4.1). This highlights how uncertainty about substitution rates for individual branches – a source of

uncertainty that is explicitly accounted for with a UCLN relaxed clock but not with a strict clock – leads to uncertainty in divergence time estimates. Further, when a UCLN relaxed clock is used, 95% HPD widths are considerably wider in three-taxon time calibrated phylogenies compared to skeletal time-calibrated phylogenies (Appendix 6, Table 4.1). This is likely to reflect the fact that in the skeletal time-calibrated phylogenies, the birth-death branching process – which makes assumptions about divergence times – causes greater precision in divergence time estimates. This contrasts to the three-taxon time-calibrated phylogenies, where the branching process does make any assumptions about divergence times. This result has concerning implications because it suggests that the precision of divergence time estimates is sensitive to the assumption of the birth-death branching process that the speciation and extinction rate are constant. Given that this assumption is often likely to be violated, this precision may be misleading. Nonetheless, I tentatively suggest that the uncertainty expressed in the three-taxon time-calibrated phylogenies may be excessive given that all the time-calibrated phylogenies inferred in treePL corroborate with each other, and with the skeletal time-calibrated phylogenies inferred in RevBayes. However, I make this assertion whilst also considering the potential sources of error outlined above.

For analyses performed with chloroplast data in RevBayes, the skeletal and three-taxon time-calibrated phylogenies were broadly congruent with those inferred in treePL with smoothing values of 100 and 10000 (Appendix 6 and 7, Table 4.1). Given that smoothing values of 100 and 10000 are likely to be more appropriate for this dataset, and corroborate inferences derived from nuclear data, this result supports previous inferences from this study about when major clades within *Ipomoea* evolved (Fig. 4.2). Like with the nuclear data, 95% HPDs were considerably wider when a UCLN relaxed clock was used, and when divergence times were inferred in three-taxon time-calibrated phylogenies. I suggest the same explanations for these patterns as in the discussion of nuclear time-calibrated phylogenies.

Taken together, these results provide tentative support that the majority of major clades within *Ipomoea* evolved in the Late Oligocene and throughout the Early Miocene, and that the Parana Clade, Coriaceous Sepal Clade, and *Ipomoea* sect. Batatas evolved in the Late Miocene. However,

these conclusions are dependent on several important variables. These include the calibration strategy used to infer the time-calibrated phylogeny for Convolvulaceae and Solanaceae, the assumptions of different methods that account for among-branch-substitution-rate-variation, and – in Bayesian analyses – the assumption of a constant rate of speciation and extinction. It is important to consider the effect of these variables when interpreting the divergence time estimates presented here. Specifically, the calibration strategy will affect inferences of the absolute age of the entire of *Ipomoea*, whilst assumptions about the magnitude and nature of among-branch-substitution-rate-variation, and the constancy of speciation and extinction rates, will affect inferences of the relative ages of different clades within *Ipomoea*.

Significant variation in diversification parameters among clades within Ipomoea is highly likely

Inferences of diversification parameters are sensitive to uncertainty and error in divergence time estimates. In order to account for this, diversification parameters were inferred on time-calibrated phylogenies inferred with different smoothing values in treePL. With the nuclear time-calibrated phylogenies, a significant net diversification rate increase was inferred in the clade that contains both the Parana Clade and Coriaceous Sepal Clade. This net diversification rate increase was strongly supported, regardless of the smoothing value (Fig. 4.3a, Appendix 9). This inferred net diversification rate increase is unsurprising. Species within the Parana Clade and Coriaceous Sepal Clade often have very localised geographical ranges, contain groups of species whose morphologies overlap significantly, but overall, exhibit a variety of different growth forms and occur in a range of different habitats. These are characteristic attributes of rapid evolutionary radiations (Hughes and Eastwood. 2006; Givnish et al. 2009; Nevado et al. 2016).

With the chloroplast time-calibrated phylogenies inferred with smoothing values of 100 and 10000, a significant net diversification rate increase was inferred in the Parana Clade but not the Coriaceous Sepal Clade. By contrast, no significant rate shifts were inferred when a smoothing value of 0.01 or 1 was used (Fig. 4.3b, Appendix 9). However, given that cross-validation analyses suggest smoothing values of 100 and 10000 are more appropriate, I suggest that the chloroplast dataset supports a significant net diversification rate increase at the origin of the Parana Clade. Taken

together, the nuclear and chloroplast time-calibrated phylogenies provide strong support for a significant net diversification rate increase at either the origin of the clade containing the Parana Clade and Coriaceous Sepal Clade, or at the origin of the Parana Clade.

Despite the likely presence of rate shifts within *Ipomoea*, there is considerable uncertainty about the nature of these shifts. First, the possibility of erroneous divergence time estimates means there is considerable uncertainty about when the rate shifts actually occurred. Although divergence time estimation error was accounted for to an extent, by inferring diversification parameters on time-calibrated phylogenies inferred with different smoothing values, other factors were not accounted for. For example, calibration uncertainty may mean that the entire phylogeny is calibrated to an incorrect timescale. A further factor that will likely affect the inferred timing and number of rate shifts, is the limited taxon sampling within the phylogeny. For example, in the Parana Clade, only around 33% of the species are sampled. Although this was corrected for in BAMM, by specifying sampling proportions, a robust and well-resolved phylogeny that includes these missing species may result in the inference of several nested rate shifts that could not be recovered in the analyses presented here.

A further aspect of these rate shifts for which considerable uncertainty remains, is whether they reflect changes in the speciation rate or extinction rate. The analyses presented here suggest they result from significant increases in the speciation rate. However, in many cases it is likely to be very difficult to determine the relative contribution of speciation rate changes and extinction rate changes on inferred diversification dynamics (Rabosky 2010; 2014; Sanderson and Donoghue 2015; Moore et al. 2016). This is primarily because extinct branches are not sampled in a molecular phylogeny. Uncertainty about the relative importance of changes in speciation rates and changes in extinction rates will have profound implications for how hypothesis are constructed to explain the net diversification rate shifts discussed here. Further, a failure to sample extinct branches may mean that other aspects of the diversification history of *Ipomoea* have been entirely overlooked in this study.

The sweet potato evolved in pre-human times

Time-calibrated phylogenies were inferred for *Ipomoea* in treePL that used a point calibration for the root node that was equal to the MPE for the age of *Ipomoea* from Calibration Strategy 4. This calibration is likely to lead to significant underestimation of divergence times within *Ipomoea***,** because the calibration strategy from which it is derived is likely to bias estimates to younger ages. Biasing the analysis in this way represents the most robust test of whether the storage root of *I. batatas* evolved in pre-human times. In these time-calibrated phylogenies, the divergence between *I. batatas* and its closest wild relative *I. trifida* was always in pre-human times. This was the case regardless of the smoothing value, or whether the phylogeny was inferred from nuclear or chloroplast data (Fig 4.4. Appendix 10). When divergence times between *I. batatas* and its closest wild relatives were also inferred in a Bayesian framework, the inferred divergence between *I. batatas* and *I. trifida* also significantly predated human times. This was the case regardless of whether a strict clock or a UCLN relaxed clock was used, or whether the phylogeny was inferred with nuclear or chloroplast data (Fig. 4.4, Appendix 10). *I. batatas* is the only species amongst its close relatives that possesses a storage root, and together these results indicate that the storage root of *I. batatas* evolved in pre-human times.

Additional analyses were performed to further query the conclusion that the divergence between *I. batatas* and *I. trifida* occurred in pre-human times. In these additional analysis, which were performed with the chloroplast dataset, the divergence between *I. batatas* and *I. trifida* was constrained such that it occurred 15 Ka. In order to satisfy this calibration, the inferred substitution rate for the plastome was approximately $2*10^{-2}$ substitutions Myr^{-1} (Fig. 4.5, Appendix 11). This is far higher than any previously recorded substitution or mutation rate for chloroplast DNA. Although there might be an expectation of elevated substitution rates for some genes in association with domestication, an increase in rates of this magnitude across the entire plastome is highly unlikely. This result therefore supports the conclusion that the divergence between *I. batatas* and *I. trifida* occurred in pre-human times.

The analyses presented in this study also show that the majority of diversity within *I. batatas* is likely to have evolved in pre-human times. In time-calibrated phylogenies constructed from all sampled accessions of *I. batatas* and *I. trifida*, divergence events between different accessions significantly predated human times (Fig. 4.6, Table 4.2, Appendix 12). This was the case regardless of the method used. These different methods included using treePL with different smoothing values, Bayesian analyses with a strict clock and UCLN relaxed clock, and Bayesian analyses with three-taxon time-calibrated phylogenies that removed temporal assumptions associated with the birth-death tree prior.

Assimilating uncertainty and making conclusions

It is challenging to assimilate the uncertainty in the results presented here. By highlighting the assumptions that underpin different methods, the different possibilities of potential uncertainty can appear immense. In most studies performed in a Bayesian framework, the 95% HPD is presented as an effective way of accounting for uncertainty. However, as demonstrated here, the 95% HPD is entirely dependent on the assumptions that are made when performing an analysis. When the validity of different sets of assumptions is difficult to distinguish, as is often the case when inferring divergence times and diversification parameters, no single 95% HPD is therefore capable of reflecting the uncertainty inherent in these different assumptions. Thus, a more valuable approach for future studies, as followed here, may be to comprehensively account for the different types of assumptions that underlie inferences about the temporal dynamics of evolutionary diversification and openly discuss the implications of these. This may not provide simple answers to questions, but may provide a more robust foundation upon which future studies of macroevolution can rest.

Despite the uncertainty presented here, if questions are carefully framed, and results carefully interpreted, novel and robust inferences about the temporal dynamics of evolutionary history can be made. In the analyses presented here, this includes the inference that the storage root of *I. batatas* evolved in pre-human times and was not a result of human domestication, and the inference that there is likely to be at least one significant shift in net diversification rates within *Ipomoea*.

Regardless, in some cases it remains difficult to perform analyses from which robust inferences and confident conclusions about the temporal dynamics of evolutionary history can be made. Even when this is the case, for example with respect to the age of *Ipomoea* or major clades within *Ipomoea*, or the timing and frequency of net diversification rate shifts, exploring the uncertainty in these inferences is invaluable. It can provide a basis upon which to develop new hypotheses that can be meaningfully tested with available data in the future, as well as providing a framework for prioritising the collection of new data. Further, it can stimulate theoretical and methodological studies that aim to characterise the extent to which inferences about macroevolutionary history can be made from analysing molecular phylogenies.

5. CONCLUSION

Here, I have characterised theoretical limits of inferences that can be made about macroevolutionary history from a robust phylogenetic framework. I have explicitly focussed on divergence time and diversification parameter estimation. These are two of the most fundamental variables in macroevolution, and together control the temporal dynamics of evolutionary diversification. Inferences of these two variables underpin other areas of macroevolutionary research that focus on explaining patterns of evolutionary diversification. In analyses presented here, I have characterised the implications of different classes of substitution rate variation when estimating divergence times in genomic scale multilocus datasets, and showed that lineage-specific rates lead to error regardless of the quantity of sequence data that is sampled. I have also explored the implications of assumptions that commonly used methods for integrating models of substitution rate variation with multiple fossil calibrations make about the fossil record. This showed that the often unrealistic nature of assumptions about the fossil record can be a further source of error in divergence time estimates. In analyses of *Ipomoea*, I then characterised and discussed the sources of error for divergence estimates in this genus, and the implications of this for estimating accurate diversification parameters.

Characterising the theoretical limits of inferences that can be made about macroevolutionary history from a robust phylogenetic framework is important because the increasing availability of large molecular datasets – often of a genomic scale – is allowing robust phylogenies to be inferred for many groups (Nevado et al. 2016; Muñoz-Rodríguez et al. 2018; 2019). Coupled with the development of complex methodologies for making inferences about macroevolution from these phylogenies, this has led to ambitious claims about the inferences that can be made about the macroevolutionary history of different groups (Lagomarsino et al. 2016; Cardillo et al. 2017; Folk et al. 2019).

In addition to investigating purely theoretical issues, I also inferred divergence times and diversification parameters within *Ipomoea*. *Ipomoea* is a large tropical genus containing the economically important Sweet Potato. Given that a robust phylogenetic and taxonomic framework has only recently been assembled for *Ipomoea* (Muñoz-Rodríguez et al 2018; 2019), studies of divergence

times and diversification parameters within *Ipomoea* provided several novel insights that make an important contribution to our understanding of tropical plant evolution, and the evolutionary context in which economically important crops evolve. Further, the investigation of divergence times and diversification parameters within *Ipomoea* explicitly accounted for the theoretical concepts that were examined in this project. Inferences about divergence times and diversification parameters, and the conclusions drawn from these inferences about the macroevolutionary history of *Ipomoea*, therefore account for the theoretical shortcomings that are inherent in many methods.

KEY FINDINGS

Lineage-specific rates lead to erroneous divergence time estimates in genomic scale datasets

As part of the investigation of theoretical issues, I first characterised the implications of substitution rate variation on divergence time estimates in a species tree when analysing genomic scale, multi-locus datasets (Carruthers et al. 2019). I evaluated the impact of lineage-specific rates – whereby among-branch-substitution-rate-variation acts consistently across the entire genome, residual rates – whereby the pattern of among-branch-substitution-rate-variation is different at each sampled locus, and gene-specific rates – whereby the average rate across all branches varies at each sampled locus. It was shown that lineage-specific rates cause erroneous divergence time estimates regardless of how many loci are sampled. This is because lineage-specific rates lead to among-branch-substitution-rate-variation in a species tree regardless of the number of sampled loci, and molecular sequence data does not provide direct information about the magnitude of these among-branch rate differences.

Alternatively, with residual rates or gene-specific rates, error in divergence time estimates is reduced to a far greater extent when a large number of loci is sampled. This is because with residual rates or gene-specific rates, there are no rate differences between branches in a species tree when a large number of loci is sampled. In contrast to this, when a large quantity of sequence data is sampled from a single locus and there are residual rates, significant error in divergence time estimates remains.

This is because when sequence data is only sampled from a single locus and there are residual rates, there are rate differences between branches in the species tree.

These results comprehensively set out the effects of substitution rate variation on divergence time estimates in genomic scale datasets, an area that has become increasingly complex in recent years as datasets have grown in size and complexity (Ho 2014; dos Reis et al 2014; Zhu et al. 2015). The fact that error resulting from lineage-specific rates is not ameliorated in genomic scale datasets is of particular importance because there is widespread evidence that lineage-specific rates are prevalent throughout much of biology, and can significantly bias divergence-time estimates (Langley and Fitch 1974; Britten 1984; Gillespie 1989, 1991, Bromham et al. 1996; Phillips and Fruciano 2018).

The use of multiple fossil calibrations with relaxed clocks does not provide a robust basis for estimating divergence times

Following investigation of the effects of rate variation, I evaluated the implications of fossil calibrations on divergence time estimates. Fossil calibrations, or occasionally other forms of external information such as a geological event, have a complex relationship with time-calibrated phylogenies. The complexity of this relationship is often misunderstood and overlooked. Primarily, a fossil calibration is required to place a phylogeny on an absolute timescale. In addition to this, when several fossil calibrations are used in a single phylogeny with a relaxed clock model, fossil calibrations will influence inferences of the temporal duration and substitution rate of individual branches (Sanderson 1997; Donoghue and Benton 2007; Magallón et al. 2015; Donoghue and Yang 2016). As such, the fossil calibrations will influence the inferred relative ages of different clades. This complex relationship means that methods that use multiple fossil calibrations with relaxed clocks make a range of important assumptions about the fossil record (Donoghue and Benton 2007). When investigating the implications of fossil calibrations for divergence time estimation, I therefore evaluated whether the assumptions of methods that use multiple fossil calibrations with relaxed clocks are realistic, and determined the extent to which methods are sensitive to these assumptions. I evaluated these points in the context of three plant clades (Osmundaceae, Convolvulaceae and Solanaceae, and Spermatophyta) for which time-calibrated-phylogenies have recently been constructed according to commonly used

methodologies (Parham et al. 2012; Särkinen et al. 2013; Grimm et al. 2015; Magallón et al. 2015). Conclusions from this study are therefore directly relevant to commonly used and recognised methods for divergence time estimation.

This study showed that the assumptions of methods that use multiple fossil calibrations with relaxed clocks are often violated. This was the case for node calibration methods, which have an implicit assumption that sets of fossil calibrations used simultaneously in a single phylogeny reflect the relative ages of different clades. In contrast to this assumption, it was shown that sets of fossil calibrations are unlikely to reflect the relative ages of different clades. It was also shown that the key assumption of the fossilised-birth-death-process (FBDP) is violated. In contrast to the assumption of the FBDP that the fossil recovery rate is constant (Stadler 2010; Gavyryushkina et al. 2014; Heath et al. 2014), analyses performed in this study presented strong evidence that instead it is highly heterogeneous. As well as demonstrating that key assumptions of methods are violated, this study also showed that methods are highly sensitive to these assumptions. This was the case for both node calibration methods and the FBDP. Further, this study also showed that analyses which use relaxed clocks are considerably more sensitive to assumptions about the fossil record in comparison to methods that use strict clocks. Taken together, the results from this chapter highlight that many divergence time estimates are likely to be unreliable, and are critically undermined by unrealistic assumptions about the fossil record.

A robust but uncertain temporal framework for the evolutionary diversification of Ipomoea

Following investigation of theoretical problems, I then inferred divergence times and diversification parameters for *Ipomoea*. In the first part of this chapter, an overall age estimate for the whole of *Ipomoea* was inferred. The most closely related fossils to *Ipomoea* that can be implemented as calibrations occur in Solanaceae (Särkinen et al. 2013; Mitchell et al. 2016; Wilf et al. 2017). To incorporate these fossils as calibrations, and infer age estimates for *Ipomoea*, time-calibrated phylogenies were constructed for Convolvulaceae (within which *Ipomoea* is nested) and its sister family Solanaceae.

When inferring time-calibrated phylogenies for Convolvulaceae and Solanaceae, several different approaches for implementing fossil calibrations were compared, referred to as calibration strategies. Different calibration strategies were used because the fossil record of Solanaceae is extremely fragmentary, and the relationship between the age of each fossil and the age of its respective clade is extremely uncertain. With different calibration strategies, age estimates for *Ipomoea* ranged from 15 – 30 Ma.

For subsequent analyses of divergence times and diversification rates throughout *Ipomoea*, an age of 30 Ma was assumed for the entire genus. However, when interpreting inferred divergence times and diversification rates, the implications of different age estimates for *Ipomoea* that derived from different calibration strategies were discussed. As well as considering the implications of different age estimates for *Ipomoea*, the implications of different assumptions about the magnitude and nature of lineage-specific rates – which analyses in Chapter 2 showed are likely to be prevalent within *Ipomoea* – were also discussed. By accounting for these different sources of uncertainty, considerable caveats surrounded conclusions about the temporal dynamics of evolutionary diversification within *Ipomoea*. Nonetheless, the results presented in this study provide an important illustration of the implications of assumptions that different methods make about the rate of molecular sequence evolution and the nature of the fossil record.

Further, regardless of the assumptions inherent in different methods, a number of robust conclusions about the temporal dynamics of evolutionary diversification within *Ipomoea* could be made. Most notably, a significant increase in net diversification rates was inferred for a clade of Neotropical *Ipomoea*. This increase in net diversification rates is robust to assumptions about the magnitude and nature of lineage-specific rates when inferring the underlying time-calibrated phylogeny, and is of a scale equivalent to some of the most iconic radiations in the plant kingdom (Baldwin and Sanderson 1998; Hughes and Eastwood 2006; Givnish et al 2009). Given species within the clade with elevated net diversification rates also show very narrow ranges and highly overlapping morphologies – two characteristics of rapid evolutionary radiations – this inferred net diversification

rate increase is likely to reflect biological reality, rather than being an artefact of flawed methodologies (Hughes and Eastwood. 2006; Givnish et al. 2008; Nevado et al. 2016).

Despite confidence in the inference of this net diversification rate increase, it is difficult to distinguish the impact of speciation and extinction rates on net diversification rates (Rabosky 2010; Sanderson and Donoghue 2015; Moore et al. 2016). As such, it is difficult to definitively conclude whether this increase in net diversification rates results from an increase in speciation rates or a decrease in extinction rates. As well as the inference of an increase in net diversification rates, a more tentative conclusion concerning the temporal dynamics of evolution in *Ipomoea* is that the majority of major clades within *Ipomoea* are likely to have evolved by the end of the Miocene.

When investigating the time of origin of the Sweet Potato, a younger estimate of 16 Ma was assumed for the age of *Ipomoea*. This younger estimate enabled a more robust test of our primary hypothesis relating to the Sweet Potato, which concerned whether the storage root of the Sweet Potato, and extant diversity within the Sweet Potato, evolved in pre-human times. The results from these analyses present conclusive evidence that the storage root of the Sweet Potato, and much of the extant diversity within the Sweet Potato, evolved in pre-human times. This contrasts markedly to the pre-existing paradigm for the origin of the storage root of the Sweet Potato, whereby it is assumed to have originated relatively recently as a result of human domestication.

PROSPECTS FOR MORE ROBUST INFERENCES IN FUTURE STUDIES

Rate Variation

A range of new and complex models of among-branch-substitution-rate-variation are being and will likely continue to be developed such as the random local clock (Drummond and Suchard 2011), the mixed relaxed clock (Lartillot et al. 2016), and new frameworks for analysing rate variation at multiple loci (dos Reis et al. 2014). It is questionable whether these methods will improve the accuracy of divergence time estimates. This is because regardless of these methods, molecular sequence data does not provide direct information about rate or time. Analyses which use new and complex methods will therefore remain sensitive to the underlying assumptions that they make about

rate and time. Although assumptions are often considerably more complex in these new methods, there is very little evidence relating to whether or not these assumptions are realistic, and the extent to which this varies in different contexts.

An alternative possibility for inferring patterns of among-branch-substitution-rate-variation more accurately, and estimating more accurate divergence times, is the further development of models that directly link patterns of rate variation to specific characters such as woodiness, generation time, or whether the plant grows in tropical or temperate conditions (Lartillot and Poujol 2011; Ho 2014; Berv and Field 2018). Several studies have been undertaken to determine the relationship between patterns of rate variation and specific characters (Gillespie 1989, 1991; Duff and Nickrent 1997; Lanfear et al. 2013; Ho 2014), and models have been developed to jointly estimate rate variation and character evolution on a phylogeny and use this as a basis for estimating divergence times (Lartillot and Poujol 2011). Models which link rate variation to specific characters may significantly improve divergence time estimates in certain contexts, particularly where the evolution of a character has been studied in depth and its relationship with rate variation is understood in detail.

However, in many cases, the evolution of a character on a phylogeny is difficult to model. This will cause fundamental problems for methods that jointly measure character evolution and substitution rates. Problems with modelling character evolution may in part result from the difficulty of appropriately defining and measuring characters (for example Nelson and Platnick 1981; Wilkson 1995; Scotland et al. 2003). Further issues will result from the inherent differences between an inferred phylogeny, and the actual macroevolutionary history of the relevant group. For example, if a phylogeny is inferred with limited taxon sampling and there are many extinct branches, the character states of many branches will be unknown. This can potentially jeopardise inferences of character evolution (Nelson and Platnick 1981; Patterson 1981; Frohlich 1987; Donoghue et al. 1989; Maddison 1990).

As well as problems that relate specifically to modelling character evolution, further problems may result from limitations in understanding of the relationship between character evolution and substitution rate variation. These problems in part result from the fact that the relationship between

character evolution and rate variation has so far only been investigated for a fairly small number of different characters and in a very limited number of different contexts (Gillespie 1989, 1991; Duff and Nickrent 1997; Lanfear et al. 2013, Berv and Field 2018). As such, there is an insufficient understanding of the relationship between character evolution and rate variation to enable robust inferences for many groups.

In addition to the limited context in which studies of the relationship between character evolution and substitution rate variation have been performed, the nature of studies that have investigated this relationship also have important limitations. Studies of this relationship typically involve the comparison of branch lengths amongst sister taxa in a molecular phylogeny (Gillespie 1989, 1991; Duff and Nickrent 1997; Lanfear et al. 2013), or are performed in the context of a previously inferred time-calibrated phylogeny (Berv and Field 2018). Analyses based on sister taxa comparisons provide a robust basis to determine the relative substitution rates of each sister taxon, and thus the relationship between patterns of rate variation and a character of interest. However, because sister taxa are necessarily extant, analyses which use this approach will likely overlook patterns of rate variation at deeper branching positions. By contrast, analyses performed in the context of a previously inferred time-calibrated phylogeny will inevitably be sensitive to the accuracy of the time-calibrated phylogeny.

Taken together, these limitations – that concern character evolution, and the relationship between character evolution and substitution rates – mean that in many cases it will be difficult to implement methods that jointly model character evolution and among-branch-substitution-rate-variation, and to assess the robustness of inferences that result from them.

Fossil calibrations

The optimal way to integrate fossil calibrations into molecular phylogenies as a basis for estimating accurate divergence times has been, and continues to be, an active area of research (Near and Sanderson 2004; Ho and Phillips 2009; Heath 2012; Warnock et al. 2012, 2015; Heath et al. 2014). Although the variety and complexity of methods for implementing fossil calibrations continues

to increase, the fundamental principle that underlies all methods is the expression of an assumption about the relationship between the ages of fossils that are implemented as calibrations and the actual ages of their respective clades. This is the case for traditional node calibration methods (Ho and Phillips 2009), more recently developed node calibration methods that use machine learning methods to derive optimal calibration densities (Heath 2012), and tip-calibration methods (including the fossilised-birth-death-process) that derive the relationship between fossil ages and clade ages more indirectly through the implementation of a "morphological clock" and/or an inferred fossil recovery rate (Ronquist et al 2012; Heath et al. 2014). The importance of assumptions about the relationship between the ages of fossils and the ages of clades is also critical in methods that are not conducted in an explicitly phylogenetic context, but which infer age estimates by modelling the entire stratigraphic ranges of different taxa (for example Silvestro et al. 2015).

In many groups, the nature of the relationship between the ages of fossils and the ages of their respective clades is poorly understood (Wilf and Escapa 2015; Wilf et al. 2017). Further, other sources of evidence which are incorporated into integrated analyses, such as large molecular datasets, do not provide further evidence about the nature of this relationship. As such, analyses which implement multiple fossil calibrations will remain sensitive to often arbitrary assumptions about the fossil record – as was the case in the analyses performed in Chapter 3.

In the future, new fossils will undoubtedly be discovered across the tree of life. This may mean that many clades will have considerably richer fossil records. This in turn may enable divergence times to be inferred with a significantly higher level of accuracy and precision – as is already likely to be the case in the limited number of model datasets for which new methods of divergence time estimation are typically evaluated (Yang and Rannala 2006; Drummond and Suchard 2011; Heath 2012; Heath et al. 2014; Lartillot et al. 2016; Gavryushkina et al. 2017).

However, for the overwhelming majority of plant clades, the fossil record is far from the completeness of the model datasets with which new methods are typically evaluated. Future discoveries are unlikely to significantly change this fact, especially given that for many plant clades the soft body parts make the precise taxonomic assignment of fossils problematic. Further, even if the

fossil record for a particular clade is markedly improved, such that its completeness is similar to that of the well-preserved fossil record of Osmundaceae, improvements in the accuracy of divergence time estimates are unlikely – as illustrated in Chapter 3. Taken together, a reliance on the future improvement of the fossil record is a problematic basis upon which to expect there to be widespread improvements to the precision and accuracy of divergence time estimates.

In the future, it is therefore likely to remain difficult to derive methods for implementing fossil calibrations that lead to reliable divergence time estimates. This will remain the case regardless of the complexity of different methods, and in the context of future fossil discoveries. As such, when divergence time estimates are presented in future studies, a practical approach will be to account for the uncertainties and assumptions inherent in different methodologies. One way in which this can be done is to perform divergence time analyses with a range of different methods, and then combine inferred divergence times into a single set of age estimates that account for different inferences from each method (Betts et al. 2018). The strength of this approach is that it integrates the uncertainty from different methods into a single set of age estimates, thus providing a single result that is relatively straightforward to interpret.

However, when divergence times are combined from different methods that make varied assumptions about clade ages in relation to fossil ages, it is harder to determine the implications of assumptions of individual methods. Further, in some cases it is also questionable whether combining divergence time estimates from different methods truly reflects a lack of knowledge about the relationship between the age of a fossil calibration and that of its respective clade. For example, Betts et al. (2018) inferred divergence times where fossil calibrations were implemented as either uniform calibration densities, or two alternative non-uniform calibration densities. They then combined different estimates into a single set of divergence times, in an attempt to infer divergence times that accounted uncertainty in the relationship between fossil ages and actual clade ages. However, this approach only accounts for uncertainty in the relationship between fossil ages and clade ages in the context of the necessarily limited number of ways in which fossil calibrations were implemented in the study.

A further inconsistency of the approach followed by Betts et al. (2018) is that the uniform calibration density that they used accurately reflects a lack of knowledge about the relationship between fossil calibrations and actual clade ages. This is because it explicitly reflects complete agnosticism about this relationship. This contrasts to their combined age estimate that is derived from a series of contradictory calibration densities – and as such is convoluted and difficult to meaningfully interpret.

Because of the problems outlined above, a more valuable approach in future studies may be to perform series of analyses with different methods for implementing calibrations, and present results from these analyses without attempting to combine them into a single set of age estimates. In general, such studies should initially infer divergence times using methods that make minimal assumptions about the relationship between fossil ages and clade ages, such as node calibrations implemented as uniform calibration densities with old maxima. Divergence times inferred in this context would reflect a lack of knowledge with respect to the relationship between fossil ages and clade ages. Subsequently, divergence times would be inferred with methods that make more precise assumptions about clade ages in relation to fossil ages, such as node calibrations implemented as non-uniform calibration densities. Although, the rationale of individual calibration strategies will remain difficult to justify, such an approach provides a basis to determine the implications of different assumptions. This approach was followed in Chapter 4.

A change in the general approach to macroevolutionary studies

In many groups, a more realistic avenue for future research will be to adopt an approach to macroevolutionary studies that makes less ambitious statements about evolutionary history and discusses in more detail the caveats that are linked to specific inferences. This was the approach that I followed in Chapter 4. In many instances, this will lead to inferences that are considerably less precise than is characteristic of many macroevolutionary studies, which make bold claims about when, how quickly, and why, a particular group evolved (Magallón et al. 2015; Lagomarsino et al. 2016; Cardillo et al. 2017; Folk et al. 2019). However, less precise inferences are likely to more accurately reflect

what is actually known about the natural world, and they are less likely to be entirely contradicted by subsequent findings.

A further focus that would also be valuable in future studies is to answer more specific questions, rather than simply attempting to reconstruct the precise evolutionary history of a particular group. By proposing more specific questions, analyses can be designed such that hypotheses are more explicitly tested and robust inferences made about a subject of specific interest. I explicitly followed such an approach in analyses relating to the origin of the sweet potato.

CONCLUDING REMARKS

Many of the analyses presented in this study highlight the difficulty of making inferences about macroevolution from a robust phylogenetic framework. The fundamental reason for this is the inherent difference between macroevolutionary history, and a robust phylogeny and molecular sequence data from which it is inferred. Specifically, molecular sequence data does not in itself provide direct information about a timescale for evolution, and a robust phylogeny provides limited and in many cases non-existent information about the rate of extinction. This remains the case regardless of the quantity of molecular sequence data that is analysed and the complexity of the analytical method that is implemented.

In the analyses and discussion of divergence times and diversification rates presented here, I make the key assumption that a branching process occurring over an absolute timescale – the topology of which is consistent with an inferred phylogeny – is an accurate or meaningful representation of evolutionary history. This is an underlying assumption of the vast majority of macroevolutionary studies, which attempt to infer macroevolutionary parameters for branching processes with topologies that are consistent with inferred phylogenies. Such a branching process may in many cases be a meaningful way to conceptualise evolutionary history, and provide a useful framework for describing temporal patterns of speciation – where branches split, and extinction – where branches terminate before the present (Yule 1925; Harvey et al. 1994; Nee et al. 1994; Drummond and Rambaut 2007; Alfaro et al. 2009; Rabosky 2014; Höhna et al. 2016).

However, this framework may in some cases be a misleading simplification of reality. For example, speciation events may not correspond precisely to branching events. Instead, speciation may occur along individual branches (Futuyma 1987), or if it does occur at branching events, peripatric speciation may occur such that only one new species is generated and the ancestral species continues in one descendent branch (Mayr 1982). Alternatively, the evolutionary process may be characterised by reticulation events where two branches fuse (for example Frajman and Oxelman 2007).

Underlying problems relating to the correspondence between speciation and branching events, there are also conceptual differences about what constitute species, and the differences between entities that are required to define species (Mayr 1942; Simpson 1951; Donoghue 1985; Rieppel 1986, 1991, 2007; de Queiroz 2007). In certain contexts, problems relating to species concepts can be bypassed if a suitably pragmatic approach is taken, and analyses are performed at a sufficiently broad scale such that species correspond to individual branches of evolutionary history. Such an approach is consistent with that of de Queiroz (2007). He describes how conflict between species concepts occurs at times immediately preceding and proceeding branching events, and that apart from this, they do not contradict each other. Instead, de Quieroz describes how all species concepts recognise individual branches of evolutionary history – or in the words of de Quieroz "*independently evolving meta-population lineages*" – as individual species. However, taking such a pragmatic approach may mean that important characteristics of evolutionary history are overlooked.

Taken together, the related problems concerning the appropriateness of a branching process for representing evolutionary history, and the suitability of different species concepts, are important because they concern whether macroevolutionary inferences made in the context of a branching process meaningfully reflect biological reality. These issues were not explicitly investigated in this thesis, but reflect a further element of uncertainty about the extent to which a robust phylogeny can be used to make meaningful inferences about macroevolutionary history.

Given the important sources of error for inferences of macroevolutionary parameters, and the broader conceptual problems that may underlie some methods, careful consideration needs to be paid to how inferences about macroevolutionary history are interpreted. This is the case even when

genomic scale datasets are analysed, and complex analytical methods are implemented. Interpretations of results must account for the fact that there are aspects of macroevolutionary history that will never be re-constructed. This was the approach I followed in Chapter 4, where I accounted for uncertainty that is inherent in different methods, and avoided making overly confident or precise statements about macroevolutionary history. This emphasis contrasts to the prevailing viewpoint of macroevolutionary research that describes how genomic scale datasets and new and complex analytical methods are continually enabling macroevolutionary history to be described with increasing detail, accuracy, and precision.

The supposed rationale of statements about the increased detail, accuracy, and precision of inferences about macroevolutionary history stem from the fact that increasingly large datasets are being analysed, and increasingly complex methodologies are being implemented (Heath et al. 2014; dos Reis et al. 2014; Grimm et al. 2015; Magallón et al. 2015). However, as shown here, generic statements that more data and more complex analyses inherently lead to more robust inferences may not be justified *per se*.

REFERENCES

Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G, and Harmon LJ. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc Natl Acad Sci USA.* 106:13410-13414.

APG. 1998. An ordinal classification for the families of flowering plants. *Ann Missouri Bot Gard.* 85:531-553.

Aris-Brosou S. and Yang Z. 2003. Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. *Mol Biol Evol.* 20:1947-1954.

Atchison GW, Nevado B, Eastwood RJ, Ortiz-Contreras N, Reynal C, Madriñán S, Filatov DA, Huches CE. 2016. Lost crops of the Incas: origins of domestication of the Andean pulse crop tarwi, *Lupinus mutabilis*. *Am J Bot.* 103:1592-1606.

Baldwin BG and Sanderson MJ. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc Natl Acad Sci USA.* 95:9402-9406.

Barba-Montoya J, dos Reis M, Schneider H, Donoghue PCJ, Yang Z. 2018. Constrianing uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous terrestrial revolution. *New Phytol.* 218:819-834.

Beaulieu JM, O'Meara BC, Crane P, Donoghue MJ. 2015. Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. *Syst Biol.* 64:869-878.

Beerling DJ and Osborne CP. 2006.The origin of the savanna biome. *Glob Change Biol*. 12:2023–2031.

Bell CD, Soltis DE, Soltis PS. 2005. The age of angiosperms: a molecular timescale without a clock. *Evolution.* 59:1245-1258.

Benton MJ. 1995. Testing the time axis of phylogenies. *Phil Trans R Soc Lond B.* 349:5-10.

Benton MJ and Donoghue PCJ. 2007. Paleontological evidence to date the tree of life. *Mol Biol Evol.* 24:26-53.

Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJ, Pisani D. 2018. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat Ecol Evol.* 2:1556-1562.

Bolker B, Butler M, Cowan P, de Vienne D, Eddelbuettel D, Holder M, Jombart T, Kembel S, Michonneau F, Orme D, O'Meara B, Paradis E, Regetz J, Zwickl D 2017. phylobase: base package for phylogenetic structures and comparative data. R package version 0.8.4. Available at: https://CRAN.R-project.org/package=phylobase.

Bomfleur B, Grimm GW, McLoughlin S. 2015. *Osmunda pulchella* sp. nov. from the Jurassic of Sweden - reconciling molecular and fossil evidence in the phylogeny of modern royal ferns (Osmundaceae). *BMC Evol Biol.* 15:126.

Britten RJ. 1984. Rates of DNA sequence evolution differ between taxonomic groups. *Science*.

39:1393–1398.

Britton T. 2005. Estimating divergence times in phylogenetic trees without a molecular clock. *Syst Biol*. 54:500–507.

Bromham L. 2006. Molecular dates for the Cambrian explosion: is the light at the end of the tunnel an oncoming train? *Palaeontol Electron.* 9:2004–2006.

Bromham L, Rambaut A, Harvey PH. 1996. Determinants of rate variation in mammalian DNA sequence evolution. *J Mol Evol.* 43:610–621.

Brown JW and Smith SA. 2018. The past sure is tense: on interpreting phylogenetic divergence time estimates. *Syst Biol*. 67:340-353.

Buffon C. 1761. Histoire naturelle générale et particulière, vol. 9. Paris.

Buffon C. 1766. Histoire naturelle générale et particulière, vol. 14. Paris.

Camin J and Sokal R. 1965. A method for deducing branching sequences in phylogeny. *Evolution*. 19:311-326.

Candolle AP. 1820. Géographie botanique. pp. 359-422 *in* Dictionnaire des sciences naturelles. Strasbourg.

Cardillo M, Weston PH, Reynolds ZKM, Olde PM, Mast AR, Lemmon EM, Lemmon AR, Bromham L. 2017. The phylogeny and biogeography of *Hakea* (Proteaceae) reveals the role of biome shifts in a continental plant radiation. *Evolution.* 71:1928-1943.

Carruthers T, Muñoz-Rodríguez P, Wood JRI, Scotland RW. 2020. The temporal dynamics of evolution in *Ipomoea*. *Mol Phylogenetics Evol.* In review.

Carruthers T, Sanderson MJ, Scotland RW. 2019. The implications of lineage-specific rates for divergence time estimation. *Syst Biol.* Advanced access.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17:540–552.

Condamine FL, Rolland J, Morlon H. 2013. Macroevolutionary perspectives to environmental change. *Ecol Lett.* 16:72-85.

Contreras-Ortiz N, Atchison GW, Hughes CE, Madriñán S. 2018. Convergent evolution of high elevation plant growth forms and geographically structured variation in Andean *Lupinus* (Fabaceae). *Bot J Linn Soc.* 187:118-136.

Crisp MD, Arroyo MTK, Cook LG, Gandolfo MA, Jordan GJ, McGlone MS, Weston PH, Westoby M, Wilf P, Linder HP. 2009. Phylogenetic biome conservatism on a global scale. *Nature.* 458:754-756.

Darwin C. 1859. On the Origin of Species. John Murray: London.

Donoghue MJ. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proc Natl Acad Sci USA.* 105:11549–11555.

Donoghue MJ, Bell CD, Li J. 2001. Phylogenetic patterns in northern hemisphere plant geography. *Int J Pl Sci*. 162(Suppl.):S41–S52.

Donoghue PCJ and Benton MJ. 2007. Rocks and clocks: calibrating the tree of life using fossils and molecules. *Trends Ecol Evol.* 22:424-431.

Donoghue PCJ and Yang Z. 2016. The evolution of methods for establishing evolutionary timescales. *Phil Trans R Soc B.* 371:20160021.

Doyle JA, Biens P, Dorenkamp A, Jardiné S. 1977. Angiosperm pollen from the pre-Albian Cretaceous of equatorial Africa. *Bull Cent Rech Explor.* 1:451-473.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLOS Biol.* 4:e88.

Drummond AJ and Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.

Drummond AJ and Suchard MA. 2011. Bayesian random local clocks, or one rate to rule them all. *BMC Biol.* 8:114.

Drummond CS, Eastwood RJ, Miotto STS, Hughes CE. 2012. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete taxon sampling. *Syst Biol.* 61:443-460.

Edwards EJ, Osborne CP, Strömberg CAE, Smith SA, C4 Grasses Consortium. 2010. The origins of C4 grasslands: integrating evolutionary and ecosystem science. *Science*. 328: 587–591.

Farris JS. 1977. Some further thoughts on Le Quesne's methods. *Syst Zool.* 26:220-223.

Farris JS and Kluge AG. 1979. A botanical clique. *Syst Zool.* 28:400-411.

Fitzjohn RG. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol Evol.* 3:1084-1092.

Folk RA, Stubbs RL, Mort ME, Cellinese N, Allen J, Soltis PS, Soltis DE, Guralnick RP. 2019. Rates of niche and phenotype evolution lag behind diversification in a temperate radiation. *Proc Natl Acad Sci USA.* 116: 10874-10882

Frajman B and Oxelman B. 2007. Reticulate phylogenetics and phytogeographical structure of *Heliosperma* (*Sileneae*, Caryophyllaceae) inferred from chloroplast and nuclear DNA sequences. *Mol Phylogenetics Evol*. 43:140-155.

Futuyma DJ. 1987. On the role of species in anagenesis. *Am Nat.* 130:465-473.

Gandolfo MA, Nixon KC, Crepet WL. 2008. Selection of fossils for calibration of molecular dating models. *Ann Missouri Bot Gard.* 95:34-42.

Gavryushkina A, Welch D, Stadler T, Drummond AJ. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLOS Comput Biol.* 10:e1003919.

Gavryushkina A, Heath TA, Ksepka DT, Stadler T, Welch D, Drummond AJ. 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Syst Biol.* 66:57-73.

Gillespie JH. 1989. Lineage effects and the index of dispersion of molecular evolution. *Mol Biol Evol.* 6:636-647.

Gillespie JH. 1991. The causes of molecular evolution. Oxford: Oxford University Press.

Givnish TJ, MIllam KC, Mast AR, Paterson TB, Theim TJ, Hipp AL, Henss JM, Smith JF, Wood KR, Sytsma KJ. 2009. Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proc R Soc B.* 276:407-416.

Grimm GW, Kapli P, Bomfleur B, McLoughlin S, Renner SS. 2015. Using more than the oldest fossils: dating Osmundaceae with three Bayesian clock approaches. *Syst Biol.* 64:396-405.

Harvey PH, May RM, Nee S. 1994. Phylogenies without fossils. *Evolution*. 48:523-529.

Heath TA. 2012. A hierarchical Bayesian model for calibrating estimates of species divergence times. *Syst Biol.* 61:793-809.

Heath TA, Holder MT, Huelsenbeck JP. 2012. A Dirichlet process prior for estimating lineage-specific substitution rates. *Mol Biol Evol*. 29:939-955.

Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proc Natl Acad Sci USA.* 111:E2957-E2966.

Hedges SB, Tao Q, Walker M, Kumar S. 2018. Accurate timetrees require accurate calibrations. *Proc Natl Acad Sci USA.* 115:E9510-9511.

Hennig W. 1966. Phylogenetic systematics. University of Illinois Press: Urbana.

Ho SYW. 2014. The changing face of the molecular evolutionary clock. *Trends Ecol Evol.* 29:496–503.

Ho SYW and Phillips MJ. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst Biol.* 58:367-380.

Höhna S. 2014. Likelihood inference of non-constant diversification rates with incomplete taxon sampling. *PLOS ONE*. 9:e84184.

Höhna S. 2015. The time-dependent reconstructed evolutionary process with a key-role for mass-extinction events. *J Theor Biol*. 380:321–331.

Höhna S, Heath TH, Boussau B, Landis MJ, Ronquist F, Huelsenbeck JP. 2014. Probabilistic graphical model representation in phylogenetics. *Syst Biol.* 63:753–771.

Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst Biol.* 65:726–736.

Höhna S, Stadler T, Ronquist F, Britton T. 2011. Inferring speciation and extinction rates under different species sampling schemes. *Mol Biol Evol*. 28:2577–2589.

Hori H. and Osawa S. 1979. Evolutionary change in 5S RNA secondary structure and a phylogenic tree of 54 5S RNA species. *Proc Natl Acad Sci USA.* 76:381-385.

Huelsenbeck JP, Nielsen R, Bollback JP. 2003. Stochastic mapping of morphological characters. *Syst Biol*. 52:131-158.

Hughes CE. and Eastwood R. 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proc Natl Acad Sci USA.* 103:10334-10339.

Hughes CE, Nyler R, Linder HP. 2015. Evolutionary plant radiations: where, when, why and how? *New Phytol.* 207:247-253.

Hughes CE, Pennington RT, Antonelli A. 2013. Neotropical plant evolution: assembling the big picture. *Bot J Linn Soc.* 171:1-18.

Huelsenbeck JP and Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics.* 17:754–755.

Humboldt. 1816. Sur les lois que l'on observe dans la distribution des forms végétales. *Ann Chimie Physique*. 1:225-39.

Jukes TH. and Cantor CR. 1969. Evolution of protein molecules. *Mammalian Protein Metabolism*. 3:21–132.

Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066.

Katoh K and Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.

Kishino H, Thorne JL, Bruno WJ. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol Biol Evol.* 18:352–361.

Koenan EJM, de Vos JM, Atchison GW, Simon MF, Schrire BD, de Sousa ER, de Queiroz LP, Hughes CE. 2013. Exploring the tempo of species diversification in legumes. *S Afr J Bot.* 89:19-30.

Koenen EJM, Clarkson JJ, Pennington TD, Chatrou LW. 2015. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytol*. 207:327-339.

Kumar S. and Hedges S. B. 1998. A molecular timescale for vertebrate evolution. *Nature.* 392:917-920.

Kumar S. 2005. Molecular clocks: four decades of evolution. *Nat Rev Genet.* 6:654:662.

Lagomarsino LP, Condamine FL, Antonelli A, Mulch A, Davis CC. 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytol.* 210:1430-1442.

Lanfear R, Ho SYW, Davies TJ, Moles AT, Aarssen L, Swenson NG, Warman L, Zanne AE, Allen AP. 2013. Taller plants have lower rates of molecular evolution. *Nat Commun.* 4:1879.

Langley CH and Fitch WM. 1974. An examination of the constancy of the rate of molecular evolution. *J Mol Evol.* 3:161-177.

Lartillot N, Phillips MJ, Ronquist F. 2016. A mixed relaxed clock model. *Phil Trans R Soc B*. 371:20150132.

Lavin M, Schrire BP, Lewis G, Pennington RT, Delgado-Salinas A, Thulin M, Hughes CE, Matos AB, Wojciechowski MF. 2004. Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. *Philos Trans Royal Soc.* 359:1509-1522.

Lee MSY and Palci A. 2015. Morphological phylogenetics in the genomic age. *Curr Biol.* 25:R922-

R929.

Lewis O. 2001. A likelihood approach to estimating phlyogeny from discrete morphological character data. *Syst Biol.* 50:913-925.

Linnaeus C. 1735. Systema naturae sive regna tria naturae. Haak: Leiden.

Linnaeus C. 1781. Selected dissertations from the Amoenitates Academicae. London.

Lipscomb DL. 1985. The eukaryotic kingdoms. *Cladistics.*1:127-140.

MacGintie HD. 1953. Fossil plants of the Florissant beds, Colorado. Institution of Washington: Washington.

Maddison WP. 1990. A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on certain branches of a phylogenetic tree. *Evolution*. 44:539–557.

Maddison WP, Midford PE, Otto SP. 2007. Estimating a binary character's effect on speciation and extinction. *Syst Biol.* 56:701-710.

Magallón S, Gomez-Acevedo S, Sanchez-Reyes LL, Hernandez-Hernandez T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* 207:437–453.

Magallón S, Hilu KW, Quandt D. 2013. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *Am J Bot.* 100:556-573.

Magallón S and Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. *Evolution.* 55:1762-1780.

Magallón S and Sanderson MJ. 2001. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. *Evolution.* 59:1653-1670.

Margoliash E. 1963. Primary structure and evolution of Cytochrome C. *Biochemistry.* 50:672-679.

Marshall C. 1990. Confidence intervals on stratigraphic ranges. *Paleobiology*. 16:1-10.

Marshall C. 2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *Am Nat.* 171:726-742.

Martin HA. 2000. Re-assignment of the affinities of the fossil pollen type *Tricolpites trioblatus* Mildenhall and Pocknall to *Wilsonia* (Convolvulaceae) and a reassessment of the ecological interpretations. *Rev Palaeobot Palynol*. 111:237-251.

Martin HA. 2001. The family Convolvulaceae in the Tertiary of Australia: evidence from pollen. *Austral J Bot.* 49:221-234.

Maurin O, Davies TJ, Burrows JE, Daru BH, Yessoufou K, Muasya AM, van der Bank M, Bond WJ. 2014. Savanna fire and the origins of the 'underground forests' of Africa. *New Phytol*. 204:201–214.

Mayr E. 1942. Systematics and the origin of species. Columbia University Press: New York.

Mayr E. 1954. Change of genetic environment and evolution. pp 157-180 *in* Huxley J, Hardy AC, Ford EB., eds. Evolution as a process. Allen & Unwin: London.

Mayr E. 1982. Speciation and macroevolution. *Evolution*. 36:1119-1132.

Metzgar JS, Skog JE, Zimmer EA, Pryer KM. 2008. The paraphyly of *Osmunda* is confirmed by phylogenetic analyses of seven plastid loci. *Syst Bot.* 33:31-36.

Mitchell TC, Williams BRM, Wood JRI, Harris DJ, Scotland RW, Carine MA. 2016. How the temperate world was colonised by bindweeds: biogeography of the Convolvuleae (Convolvulaceae). *BMC Evol Biol.* 16:16.

Miyata T, Yasunaga T, Nishida T. 1980. Nucleotide sequence divergence and functional constraint in mRNA evolution. *Proc Natl Acad Sci USA.* 77:7328-7332.

Moore BR, Höhna S, May MR, Rannala B, Huelsenbeck JP. 2016. Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. *Proc Natl Acad Sci USA.* 113:9569-9574.

Morlon H, Lewitus E, Condamine FL, Manceau M, Clavel J, Drury J. 2016. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Methods Ecol Evol*. 7:589-597.

Morlon H, Parsons TL, Plotkin JB. 2011. Reconciling molecular phylogenies with the fossil record. *Proc Natl Acad Sci USA*. 108:16327-16332.

Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. 2018. The timescale of early land plant evolution. *Proc Natl Acad Sci USA.* 115:E2274-E2283.

Muñoz-Rodríguez P, Carruthers T, Wood JRI, Williams BRM, Weitemier K, Kronmiller B, Ellis D, Anglin NL, Longway L, Harris SA, Rausher MD, Kelly S, Liston A, Scotland RW. 2018. Reconciling conflicting phylogenies in the origin of sweet potato and dispersal to polynesia. *Curr Biol*. 28:1246-1256.

Muñoz-Rodríguez P, Carruthers T, Wood JRI, Williams BRM, Weitemier K, Kronmiller B, Goodwin Z, Sumadijaya A, Anglin NL, Filer D, Harris D, Rausher MD, Kelly S, Liston A, Scotland RW. 2019. A taxonomic monograph of Ipomoea integrated across phylogenetic scales. *Nat Plants.* 5:1136-1144.

Near TJ, Meylan PA, Shaffer HB. 2005. Assesing concordance of fossil calibration points in molecular clock studies: an example using turtles. *Am Nat.* 165:137-146.

Near TJ and Sanderson MJ. 2004. Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. *Phil Trans R Soc Lond B.* 359:1477-1483.

Nee S. 2006. Birth-death models in macroevolution. *Annu Rev Ecol Evol Syst.* 37:1-17.

Nee S, May RM, PH. Harvey. 1994. The reconstructed evolutionary process. *Phil Trans R Soc B.* 344:305-311.

Nee S, Mooers AO, Harvey PH. 1992. Tempo and mode of evolution revealed from molecular phylogenies. *Proc Natl Acad Sci USA.* 89:8322-8326.

Nelson G. and Platnick N. 1981. Systematics and biogeography. Columbia University Press: New York.

Nevado B, Atchison GW, Hughes CE, Filatov DA. 2016. Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. *Nat Commun.* 7:1-9.

Nürk NM, Michling F, Linder HP. 2018. Are the radiations of temperate lineages in tropical alpine ecosystems pre-adapted? *Glob Ecol Biogeogr.* 27:334-345.

Nürk NM, Atchison GW, Hughes CE. 2019. Island woodiness underpins accelerated disparification in plant radiations. *New Phytol.* 224:518-531.

Pagel MD. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc R Soc Lond B.* 255:37–45.

Pagel MD. 1999. Inferring the historical patterns of biological evolution. *Nature*. 401:877–884.

Parham JF, Donoghue PCJ, Bell CJ, Calway TD, Head JJ, Holroyd PA, Inoue JG, Irmis RB, Joyce WG, Ksepka DT, Patané JSL, Smith ND, Tarver JE, van Tuinen M, Yang Z, Angielczyk KD, Greenwood JM, Hipsley CA, Jacobs L, Makovicky PJ, Müller J, Smith KT, Theodor JM, Warnock RCM, Benton MJ. 2012. Best practices for justifying fossil calibrations. *Syst Biol.* 61:346-359.

Pennington RT and Hughes CE. 2014. The remarkable congruence of new and old world savanna origins. *New Phytol*. 204:4-6.

Pennington RT, Lavin M, Oliveira-Filho A. 2009. Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. 40:437-457. *Annu Rev Ecol Evol. Syst.* 40:437-457.

Pennington RT, Lavin M, Prado DE, Pendry CA, Pell SK, Butterworth CA. 2004. Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. *Proc R Soc B.* 359:515-537.

Phillips MJ and Fruciano C. 2018. The soft explosive model of placental mammal evolution. *BMC Evol Biol.* 18:104.

de Queiroz K. 2007. Species concepts and species delimitation. *Syst Biol*. 56:879-886.

Rabosky DL. 2010. Extinction rates should not be estimated from molecular phylogenies. *Evolution*. 64:1816-1824.

Rabosky DL. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* 9:e89543.

Rabosky DL and Lovette IJ. 2008. Density-dependent diversification in North American wood warblers. *Proc R Soc B.* 275:2363-2371.

Raup DM. 1972. Taxonomic diversity during the Phanerozoic. *Science*. 177:1065-1071.

Ree RH, Moore BR, Webb CO, Donoghue MJ, Crandall K. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution*. 59:2299–2311.

dos Reis M. and Yang Z. 2013. The unbearable uncertainty of Bayesian divergence time estimation. *J Syst Evol.* 51:30-43.

dos Reis M, Zhu T, Yang Z. 2014. The impact of the rate prior on Bayesian estimation of divergence times with multiple loci. *Syst Biol.* 63:555–565.

Renner SS. 2004. Multiple Miocene Melastomataceae dispersal between Madagascar, Africa and India. *Phil Trans R Soc Lond B*. 359:1485-1494.

Renner SS, Grimm GW, Kapli P, Denk T. 2016. Species relationships and divergence times in

beeches: new insights from the inclusion of 53 young and old fossils in a birth – death clock model. *Phil Trans R Soc B.* 371:20150135.

Ridley M. 1983. The explanation of organic diversity: the comparative method and adaptations for mating. Oxford Univ. Press: Oxford.

Rieppel O. 1988. Fundamentals of comparative biology. Birkhäuser: Basel.

Rieppel O. 1991. Things, taxa, and relationships. *Cladistics*. 7:93-100.

Rieppel O. 2007. Species: kinds of individuals or individuals of a kind. *Cladistics*. 23:373-384.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liang L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61:539–542.

Rieppel O. 1986. Species are individuals. *Evol Biol*. 20:283-317.

Rubinstein CV, Gerrienne P, de la Puente GS, Astini A, Steemans P. 2010. Early Middle Orodovician evidence for land plants in Argentina (eastern Gondwana). *New Phytol.* 188:365-369.

Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol Biol Evol.* 14:1218–1231.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol.* 19:101–109.

Särkinen T, Pennington RT, Lavin M, Simon MF, Hughes CE. 2012. Evolutionary islands in the andes: persistence and isolation explain high endemism in Andean dry tropical forests. *J Biogeogr.* 39:884-900.

Särkinen T, Bohs L, Olmstead RG, Knapp S. 2013. A phylogenetic framework for evoluitonary study of the nightshades (Solanaceae): a dated 1000 tip tree. *BMC Evol Biol.* 13:214.

Schliep K. P. 2011. phangorn: Phylogenetic analysis in R. *Bioinformatics.* 27:592–593.

Schliep K. P. 2018. phangorn: Phylogenetic reconstruction and analysis. R package Version 2.4.0. Available at: https://CRAN.R-project.org/package=phangorn.

Schuettpelz E and Pryer KM. 2009. Evidence for a Cenozoic radiation of ferns in an angiosperm-dominated canopy. *Proc Natl Acad Sci USA.* 106:11200-11205.

Scotland RW, Olmstead RG, Bennet JR. 2003. Phylogeny reconstruction: the role of morphology. *Syst Biol*. 52:539-548.

Signor PW and Lipps JH. 1982. Sampling bias, gradual extinction patterns and catastrophes in the fossil record. pp. 291-296. *in* Silver LT and Schultz PH., eds. Geological implications of impacts of large asteriods and comets on the Earth. Geol Soc Am:Boulder.

Simon MF, Grether R, de Queiroz LP, Skema C, Pennington RT, Hughes CE. 2009. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proc Natl Acad Sci USA.* 106:20359-20364.

Simon MF and Pennington T. 2012. Evidence for adaptation to fire regimes in the tropical savannas of the Brazilian cerrado. *Int J Plant Sci*. 173:711-723.

Simpson GG. 1951. The species concept. *Evolution*. 5:285-298.

Smith AB. 2001. Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Phil Trans R Soc B.* 356:351-367.

Smith AB. and McGowan AJ. 2005. Cyclicity in the fossil record mirrors rock outcrop area. *Biol Lett.* 1:443–445.

Smith SA, Beaulieu JM, Donoghue MJ. 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc Natl Acad Sci USA*. 107:5897-5902.

Smith SA and O'Meara BC. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics.* 28:2689–2690.

Sober E. 1988. Reconstructing the past. The MIT Press: Cambridge.

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS. 2000. Angiosperm phylogeny inferred from *18S* rDNA, *rbcL* and *atpB* sequences. *Bot J Linn Soc*. 133:381-461.

Spriggs EL, Clement WL, Sweeney PW, Madriñán S, Edwards EJ, Donoghue MJ. 2015. Temperate radiations and dyning embers of a tropical past: the diversification of *Viburnum*. *New Phytol.* 207:340-354.

Springer MS. 1995. Molecular clocks and the incompleteness of the fossil record. *J Mol Evol.* 41:531-538.

Srivastava G, Mehrotra RC, Dilcher DL. 2018. Paleocene *Ipomoea* (Convolvulaceae) from India with implications for an east Gondwana origin of Convolvulaceae. *Proc Natl Acad Sci USA.* 115:6028–6033.

Stadler T, Gavyryushkina A, Warnock RCM, Drummond AJ, Heath TA. 2018. The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *J Theor Biol*. 447:41-55.

Stadler T. 2009. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *J Theor Biol.* 261:58-66.

Stadler T. 2010. Sampling-through-time in birth-death trees. *J Theor Biol.* 267:396-404.

Stadler T. 2011a. Mammalian phylogeny reveals recent diversification rate shifts. *Proc Natl Acad Sci USA*. 108:6187–6192.

Stadler T. 2011b. Inferring speciation and extinction processes from extant species data. *Proc Natl Acad Sci USA.* 27:16145-16146.

Talavera G and Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56:564–577.

Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. 2012. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci USA.* 109:19333-19338.

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*. 17:57–86.

Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol.* 15:1647–1657.

Tiffney BH. 1985. The Eocene north Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. *J Arnold Arboretum, Harvard Univ*. 66:243-273.

Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics.* 27:171–180.

Warnock RCM, Parham JF, Joyce WG, Tyler R, Donoghue PCJ. 2015. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proc R Soc B.* 282:20141013.

Warnock RCM, Yang ZH, Donoghue PCJ. 2011. Exploring uncertainty in the calibration of the molecular clock. *Biol Lett.* 8:156 – 159.

Warnock RCM, Yang Z, Donoghue PCJ. 2017. Testing the molecular clock using mechanistic models of fossil preservation and molecular evolution. *Proc R Soc B.* 284:20170227.

Welch JJ. and Bromham L. 2005. Molecular dating when rates vary. *Trends Ecol Evol.* 20:320–327.

Wilf P, Carvalho MR, Gandalfo MA, Ruben Cuneo N. 2017. Eocene lantern fruits from Gondwanan Patagonia and the early origins of Solanaceae. *Science.* 355:71–75.

Wilf P. and Escapa IH. 2015. Green Web or megabiased clock? Plant fossils from Gondwanan Patagonia speak on evolutionary radiations. *New Phytol.* 207:283–290.

Wilson EO. 1965. A consistency test for phylogenies based on contemporaneous species. *Syst Zool*. 14:214–20.

Woese CR and Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA*. 74:5088-5090.

Yang Z. and Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain monte carlo method. *Mol Biol Evol.* 14:717–724.

Yang Z. and Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol* 23:212–226.

Yule GU. 1925. A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS. *Phil Trans R Soc Lond B*. 213:21–87.

Zhang C, Stadler T, Klopfstein S, Heath TA, Ronquist F. 2016. Total-evidence dating under the fossilised birth-death process. *Syst Biol.* 65:228-249.

Zhu T, dos Reis M, Yang Z. 2015. Characterization of the uncertainty of divergence time estimation under relaxed molecular clock models using multiple loci. *Syst Biol.* 64:267–280.

Zuckerkandl E. and Pauling LB. 1962. Molecular disease, evolution, and genetic heterogeneity. pp. 189–225. *in* Kasha M. and Pullman B., eds. Horizons in biochemistry. Academic Press: New York.

Zuckerkandl E. and Pauling L. 1965. Evolutionary divergence and convergence. pp. 97–166 *in* Bryson V. and Vogel HJ., eds. Evolving genes and proteins. Academic Press: New York

APPENDICES

APPENDIX 1

Table A1.1. GenBank accession numbers of sampled species for the CS dataset

Taxon	Marker					
	ITS	***trnL-trnF***	***matK***	***rbcL***	***atpB***	***ndhF***
Acnistus arborescens	DQ314183.1	KU568472.1	KU568472.1	KU568472.1	KU568472.1	KU568472.1
Anisodus luridus	AY478400.1	EU580955.1	EF438901.1	HQ216115.1	HQ215952.1	EU580856.1
Breweria rotundifolia		AY101123.1		AY101014.1	AY100805.1	
Browalia americana			EF439050.1		AM233358.1	
Browalia eludens	KP100283.1	EU580964.1				EU580865.1
Browalia speciosa				AY206719.1		
Capsicum baccatum	AF244708.1	KR078314.1	KR078314.1	KR078314.1	KR078314.1	KR078314.1
Capsicum lycianthoides	DQ314158.1	NC_026551.1	NC_026551.1	NC_026551.1	NC_026551.1	NC_026551.1
Cardioclamys madagascariensis		AY101161.1		AY101052.1	AY100840.1	
Cestrum strigalatum		EU580976.1				EU126008.1
Cestrum diurnum	JQ230969.1		GU135071	GU135235.1		
Cestrum parqui					AM233353.1	
Convolvulus canariensis	AH013855.2		KC529025.1	KC529184.1		
Convolvulus spinosus	KC528988.1		KC529123.1	KC529278.1		
Convolvulus virgatus	KC528882.1		KC529135.1	KC529290.1		
Cordisepalum phalanthopetalum		AY101159.1		AY101050.1	AY100838.1	
Dinetus racemosus		HQ412975		HQ384920.1		HQ384856.1
Duckeodendron cestroides	AY206733.1	AY206725.1		AF206763.1	AF209579.1	AY206743.1
Erycibe helwigii		AY101156.1		AY101047.1	AY100836.1	
Erycibe obtusifolia	KP093238.1		KP094013.1	KP094963.1		
Erycibe tomentosa		KU853219.1	KR024905.1	KU853163.1		
Evolvulus glomeratus		AY101121.1		AY101012.1	AY100803.1	AY936341.1
Evolvulus nuttallianus	DQ219869.1	AY101122.1	KT176618.1	KT178137	AY100804.1	

Goetzea elegans	AY206736.1	HQ412971.1		AF035738.1	HQ384772.1	AY206746.1
Hildebrandtia africana		AY101116.1	KC963387.1	AY101007.1	AY100798.1	
Humbertia madagascariensis	EU330329.1	HQ412976.1	EU330288.1	AY101062.1	HQ384777.1	HQ384857.1
Iochroma australe	DQ314189.1	EU580999.1	EF438902.1		AM233343.1	EU580892.1
Ipomoea alba	JRIW_27755					
Ipomoea biflora	AH_302 2					
Ipomoea pescaprae	*PW_464*					
Ipomoea pestigridis	*X_20100015*					
Iseia luxurians	*KP261945.1*	AY101110.1	KR024939.1	AY101001.1	AY100792.1	
Jacquemontia blanchetii	*DQ219857.1*			AY101039.1	AY100828.1	
Jacquemontia pentantha		AY101036.1		AY101036.1	AY100825.1	HQ384854.1
Jaltomata dentata					AM233341.1	
Jaltomata procumbens	*AF244710.1*	AY098695.1	KC146605.1			U47429.1
Jaltomata repandidentata				JQ594126.1		
Lycianthes heteroclita	*AF244709.1*	DQ180414.1	JQ589231.1	JQ594129.1		U72756.1
Lycium andersonii	*DQ124620.1*	DQ124562.1	AB036628.1	AB051024.1		
Lycium chinense	KU314471.1	DQ124576.1	AB036637.1	JF942339.1	FJ914152.1	FJ914028.1
Lycium cooperi	AF238984.1	DQ124579.1				
Markea panamensis	KT366066.1	EU581027.1	KJ593940.1	KJ594345.1		EU580885.1
Merremia hirta	KP261996.1	KP261996.1	KR024965.1			
Merremia quinquefolia		KF242501.1	KF242501.1	KF242501.1	KF242501.1	KF242501.1
Nicotiana glauca	AJ492410.1		AB039987.1	JQ412399.1	AM233332.1	AJ585910.1
Nicotiana rustica	AJ492415.1		AB039992.1		AM233327.1	AJ585935.1
Nierembergia hippomanica		EU581033.1				EU580917.1
Nierembergia frutescens			EF438884.1		AM233331.1	
Nierembergia scorparia				FJ911662.2		
Nierembergia linariifolia	AY560045.1					
Nolana albescens	AB019952.1/	AB036618.1	AB036647.1	AB051021.1		EU742309.1

	AB019292.1					
Nolana humifusa	AB019969.1/ AB019309.1		EF438984.1		AF209638.1	FJ914024.1
Petunia axillaris	AJ492460.1	AY098702.1	AJ585880.1	HQ384915.1		U08926.1
Petunia sp.					AM233152.1	
Physalis alkekengi	AY665850.1	DQ180420.1	AM503827.2	U08617.1	AM233321.1	U08927.1
Physochlaina physaloides	KT891913.1		EF439017.1	HQ216141.1	HQ215980.1	HQ216106.1
Porana commixta		AY101164.1		AY101055.1	AY100843.1	
Porana paniculata		AY101160.1		AY101051.1	AY100839.1	
Salpiglossis sinuata	KP100299.1	AY206730.1	EF439055.1	U08618.1	AM233297.1	U08928.1
Seddera arabica		AY101120.1	KC963389.1	AY101011.1	AY100802.1	
Solanum bulbocastanum	GQ221564.1	NC_007943.1	NC_007943.1	NC_007943.1	NC_007943.1	AF500804.1
Solanum jamesii			EF439047.1		AM233122.1	
Solanum rostratum	GQ143670.1	GQ149755.1	KT176603.1	KT178123.1	KT176261.1	U47424.1
Solanum tuberosum	KF022370.1	FJ490824.1	FN668841.1	HF572814.1	AM233351.1	L76287.1
Tetralocularia pennellii		AY101112.1		AY101003.1	AY100794.1	
Tridynamia megalantha		AY101163.1		AY101054.1	AY100842.1	

Table A1.2. GenBank accession numbers of sampled species for the Spermatophyta dataset

Species	**Marker**				
	rbcL	**atpB**	**matK**	**18S**	**26S**
Acalypha californica	AY380341	AY788199	EF135499	AY674572	EF135605
Acorus gramineus	D28866.1	AF197616.1	DQ182341.1	AF197584.1	AF036490.1
Actinidia chinensis	L01882.2	AJ235382.2	U61324.1	AF419792.1	
Actinidia arguta					AY727964.1
Aesculus pavia	U39277.2	AF035894.1		AF206838.1	AF479138.1
Aesculus glabra			AY968671.1		
Aextoxicon punctatum	X83986	AJ235384.2	DQ182342.1	AF206839.1	AF389239.1
Ailanthus altissima	U02726	AF035895	EF489111	AF206842	
Akebia quinata	L12627.2	AF209523.1	AB069851.1	L37905.1	AF389253.1
Alangium chinense	L11209.2			AF206843.1	AY260009.1
Alangium sp.		AJ235386.2			
Alangium platanifolium			U96880.1		
Alisma plantagoaquatica	L08759.1	DQ007417		AF197585.1	DQ008651.1
Alisma canaliculatum			AB040179.1		
Alnus incana	X56618.1				
Alnus sinuata		AY147101.1		AY147101.1	

Alnus japonica			AB038176.1		
Alnus glutinosa					AF479106.1
Amborella	L12628.2	AJ235389.1	AF543721.1	U42497.1	AF479238.1
Androsace erecta	AF395004.1				
Androsace sp.		AF213775.1			
Androsace sempervivoides			AY647535.1		
Androsace spinulifera				AF206847.1	AF479150.1
Androstachys johnsonii	AF206734.1	AF209527.1	AY552461.1	AF206848.1	AF479123.1
Anemopsis californica	AF197597	AF197608.1	DQ882198.1	AF197576.1	DQ008639.1
Anigozanthos flavidus	AJ404843.1	AF387600.1	AB088796.1	AF069214.1	AF290587.1
Anisophyllea fallax	AY935742	AY935849	AY935923	AY929365	AY935807
Annona muricata	AY743440.1	AJ235393.2	AF543722.1	AF206850.1	DQ008634.2
Antirrhinum majus	L11688	AJ235395	AJ429342.1	AJ236047	AY423077.1
Apium graveolens	L01885.2	AJ235396.2	AJ429370.1	AF206852.1	AF479195.1
Arabidopsis thaliana	U91966.1	NC_000932	AF144348	X16077	X52320.1
Aralidium pinnatifidum	AF299087		U58627.1		AY189036
Aristolochia macrophylla	L12630.2	AF528845.1	AB060742.1	AF206855.1	AY095450.1
Asteropeia micraster	AF206737.1	AF209533.1	AY042549.1	AF206857.1	AF479090.1
Aucuba japonica	L11210.2	AJ235402.2	AJ429318.1	U42522.1	AY727931.1
Averrhoa carambola	L14692.2	AJ235404.2	AY935924.1	AF206859.1	AF479127.1
Azorella selago	AY188417.1		AF271762.1		AY189046.1
Azorella caespitosa		GQ983635		GQ983561	
Bacopa eisenii			EF467900.1		
Bacopa monnieri	KJ773301.1				KP844741.1
Barleria prionitis	L01886	AJ236179		AF107567	AF479166
Barleria rotundifolia			JF270654.1		
Begonia sanguinea	L01888.2				
Begonia metallicaxsanguinea		AF209541.1			AF479109.1
Begonia oxyloba			AY968445		
Begonia luxurians				AF534762.1	
Berberidopsis corallina	AJ235773.1	AJ235409.2	EU002171.1	AF206866.1	AF389242.1
Bischofia javanica	AY663571.1	AY830200.1	EF135508.1	AB233605.1	EF135606.1
Boehmeria nivea	AJ235801			AF206870	AY686767
Boehmeria platanifolia			AF353579		
Bombax buonopozense	AF022118.1		AY321171.1		
Bombax ceiba		AJ233051.1		U42507.1	AF479134.1
Boopis anthemoides	L13860.1				
Boopis graminea		AJ236199.1	AJ429382.1	AF107583.1	AF479184.1
Bougainvillea glabra	M88340.1	AJ235415.2		AF206873.1	
Bougainvillea sp.			AY042560		
Bougainvillea spectabilis					HQ843443
Brasenia schreberi	M77031.1	AF209544.1	DQ185529.1	AF206874.1	DQ008661.2
Brassica napus	AF267640.1	AF267641.1	AB354273.1		
Brassica oleracea				AF513990.1	
Brassica rapa					EF470522.1
Brexia madagascariensis	L11176.1	AJ235419.2	AY935899.1	U42543.1	AF479112.1
Bursera inaguensis	L01890	AF035899		AF206877	AY177421
Bursera fagaroides			AY594462.1		
Byrsonima crassifolia	L01892	AY788206	AF344535	AY674579	EU002150
Callicarpa dichotoma	L14393	AF209551		AJ236048	
Callicarpa stapfii			FM163265		
Camellia sinensis	AF380037.1	AY725933.1	AF380077.1	AB120309.1	AY727975.1
Campanula ramosa	L13861.1	AJ235423.2			

Campanula elatines			AJ430387.1		
Campanula ramulosa				U42510.1	
Campanula trachelium					AF479191.1
Campsis radicans	AF102642	AJ236168	AF531775	AF107578	
Campynema linearis	Z77264	AJ417573	JN417414.1	GQ497570.1	AF364029
Canella winterana	AY572265.1	AF528847.1	DQ882240.1	AF206879.1	AY095455.1
Canna indica	AF378763.1	AF168892.1	AM114724.1	D29784.1	
Canna flaccida					AF205521.1
Carallia sp.	AF206744				
Carallia brachiata		AJ235425	EF135513	FJ707525.1	
Carludovica palmata	AF197596.1	AY465545.1	AB088793.1	AF293756.1	DQ008648.1
Casearia sylvestris	AF206746.1	AF209557.1		AF206882.1	
Casearia javitensis			AY935927.1		AY935809.1
Casuarina cunninghamiana	X69528.1				
Casuarina litorea		AJ235427.2			
Casuarina cristata			AY191698.1		
Casuarina equisetifolia				U42515	
Catalpa sp.	L11679.1				
Catalpa aff. speciosa		HQ384724	HQ384519		
Catalpa bignonioides				AF107579	
Caulophyllum thalictroides	AF190442.1	AF092108.1	AB069831.1	L54064.1	AF389240.1
Ceanothus sanguineus	U06795.1	AF209558.1		U42799.1	AF479102.1
Ceanothus pumilus			AF049841.1		
Celastrus scandens	AY788195	AY788264			AF222358.1
Celastrus orbiculatus			EF135517	AY788162	
Celosia argentea	AY270072.1	AF209559.1		AF206883.1	
Celosia trigyna			AY514811.1		
Celosia cristata					HQ843444.1
Celtis yunnanensi	L12638.2			U42818.1	AF479098.1
Celtis philippensis		AY263961.1	AY263925.1		
Centrolepis strigosa	AJ286558.1	AJ419131.1	DQ257507.2		AF466388.1
Cephalotus follicularis	L01894	AY788265	FJ670045.1	U42516.1	
Ceratophyllum demersum	D89473.1	AJ235430.2	AF543732.1	U42517.1	AY095456.1
Cercidiphyllum japonicum	L11673.1	AF092112.1	AM396508.1	AF094534.1	AF274639.1
Chamaedorea seifrizii	AF206748.1	AF233083.1		AF069210.1	
Chamaedorea oblongata			DQ178683.1		
Chloranthus japonicus	L12640.2	AJ235431.2			AF479245.1
Chloranthus brachystachys			AF543733.1		
Chloranthus spicatus				D29787.1	
Choristylis rhamnoides	AJ238132.1	AF274676.1	AF274609.1	AF274601.1	AF274640.1
Chrysobalanus icaco	GQ424476	AF209562	EF135519	U42519	AF479119
Chrysolepis sempervirens	AF061995.1	AF209563.1		AF209563.1	AF479107.1
Chrysolepis chrysophylla			FJ185045.1		
Cissampelos pareira	AF197590.1	AF197613.1	AJ966802.1	AF293758.1	DQ008616.1
Citrus paradisi	AJ238407.1	AJ238408.1			
Citrus trifoliata			HM163960.1		
Citrus aurantium				U38312.1	AY177420.1
Clarkia xantiana	L01896	AF209564		U67930	AF479148
Clethra alnifolia	L12609.2		AJ429281.1	AF419793.1	

Clethra arborea		AF420965.1			
Clethra cf. ferruginea					AY727968.1
Cobaea scandens	Z83143.1	AJ235440.2	L48568.1	L49277.1	AY727944.1
Coccinia sessilifolia	AY968520	AY968427	AY968446	AY973011	AY968404
Cocculus trilobus	L12642.2	AF197614.1		AF197581.1	DQ008617.1
Cocculus laurifolius			AF542588.2		
Colchicum speciosum	L12673.2	AF209569.1	AB040181.2		
Colchicum autumnale				U42072.1	AF331971.1
Connarus conchocarpus	U06798.1		EU002174.1		
Connarus championii		AY935852.1		AY929368.1	AY935810.1
Crassula marnierana	L01899.1	AJ235447.2	AF115600.1	U42525.1	AF479210.1
Cratoxylum cochinchinense	AB233891.1	AB233683.1	HQ331587.1	AB233579.1	
Crinodendron patagua	AF291940.1				AY935811.1
Crinodendron hookerianum		AF209570.1	AY491655.1	AF206893.1	
Crinum asiaticum	HM640488.1	JQ273605.1		HM640718.1	
Crinum moorei			AB017279.1		
Crinum americanum					AF293854.1
Croomia pauciflora			AY437815.1	AF168835.1	DQ008647.1
Croomia japonica	AF307493.1	AF308039.1			
Crossoma bigelovii	AY101844.1		DQ443456.1	AF193942.1	
Crossoma californicum		AF209571.1			AF479131.1
Cryptocarya obovata	L28950.1				
Cryptocarya meissneriana		AF197602.1		AF293757.1	DQ008627.1
Cryptocarya alba			AJ247158.1		
Cupaniopsis anacardioides	L13182.2	AF035903.1	AY724283.1	AF206896.1	AF479139.1
Cycas micronesica	EU016864.1	EU016826.1			
Cycas siamensis			AF410165.1		
Cyperus alternifolius	AM999802.1		AY952421.1	AY952404.1	
Cyperus albostriatus		AF168906.1			
Cypripedium calceolus	AB176549.1	AJ235448.2	AY557208.1	AF069208.1	
Cypripedium kentuckiensis					AF205119.1
Dalechampia spathulata	AY788172	AY788216	EF135525	AY788149	EF135609.1
Dampiera spicigera	X87383	AJ318971			
Dampiera diversifolia			GQ983658		
Daphniphyllum sp.	L01900.2	AF092118.1	AF274612.1	U42531.1	AF479215.1
Datisca cannabina	L21939.1	AJ235450.2	AB016467.1	AF008952.1	
Datisca glomerata					AY968411.1
Davidsonia pruriens	AF291934.1	AF209574.1	U92846.1	AF206897.1	AY935812.1
Delarbrea michieana	U50243.1	AJ236211.1	U58608.1	AF107573.1	AF479196.1
Desfontainia spinosa	Z29670.1	AJ419677.1	AJ429363.1	GQ983565.1	GQ983582.1
Didymeles perrieri	AF061994.1	AF092119.1	DQ401354.1	AF094541.1	AF389247.1
Diervilla sessilifolia	AF446937	GQ983617	AF446907	GQ983566	GQ983583
Dillenia indica	L01903.2				
Dillenia retusa		AF095732.1			AF479096.1
Dipelta floribunda	AJ420876				
Dipelta yunnanensis		GQ983629	AF446910.1	GQ983567	GQ983584
Donatia fascicularis	AF307913.1		AJ429384.1		
Donatia sp.		AJ236203.1		AJ236012.1	AF479189.1

Drosera capensis	L01909.2	AY096110.1	AY096122	U42532.1	AF389248.1
Drypetes roxburghii	M95757			U42534	
Drypetes madagascariensis		AY830256	AY552457		
Dudleya viscida	L11182.2	AJ235461.2	AF274614.1	U42526.1	AF274646.1
Elaeagnus angustifolia	U17038.1				
Elaeagnus sp.		AJ235462.2			
Elaeagnus umbellata			AY257529.1	L24090.1	
Elaeagnus bockii					JF317385.1
Elatine triandra	AY380349	AY788219	EF135532	AY674594	
Endospermum moluccanum	AJ402950	AY788220	EF135533	AY674595	EF135610.1
Erythrina cristagalli	Z70170		AY386869	AF525296	
Erythrina sousae		EU717512			
Erythrospermum phytolaccoides	AJ418798	FJ669964	EF135535	FJ669684.1	
Escallonia calcottiae	AJ419693				
Escallonia rubra		AJ318974	AJ429365		
Escallonia coquimbensis				U42544	GQ983585
Eucommia ulmoides	L01917	HQ384789	AJ429317	HQ384682	
Eucryphia lucida	L01918		EU002176.1	U42533.1	AF036494.1
Eucryphia milliganii		AJ235470			
Euphorbia polychroma	AY794827	AJ235472	EF135539		AF479125
Euphorbia pulcherrima				L37582	
Eupomatia bennettii	L12644.2	AJ235473.1	DQ401341.1	AF469771.1	DQ008636.1
Euptelea polyandra	L12645.2	AF528850.1	DQ401348.1	L75831.1	AF389249.1
Exacum affine	L11684	AJ236195	FJ014087	AJ236023	AF479180
Fagopyrum statice	AB000317.1				
Fagopyrum esculentum		EU254477.1			
Fagus grandifolia	AY263936.1	AY147105.1		AF206910.1	AY935813.1
Fagus sylvatica			AB046507.1		
Fendlera rupicola	AF206766.1	AJ236234.1	AY254063.1	AJ235986.1	AY260041.
Flacourtia indica	GU135218	AB233725	GU135055	AB233621	
Floerkea proserpinacoides	L12679.2	AF035904.1	EU002178.1	U42784.1	AF479143.1
Gaiadendron punctatum	L26072.1		DQ787445.1	L24143.1	DQ790209.1
Galax aphylla	Z80184.1				
Galax urceolata		AY725936.1	L48576.1	L49281.1	AY727983.1
Gentiana procera	L14398				
Gentiana saponaria		JF298870			JF321128
Gentiana purpurea			AJ429323		
Gentiana asclepiadea				HQ448773	
Ginkgo biloba	DQ069500.1	AJ235481.1	DQ069584.1	D16448.1	AY095475.1
Gnetum gnemon	AY296534.1	AF187060.1		U42416.1	AF036488.1
Gnetum parvifolium			AF280995.1		
Gonocaryum litorale	AJ235779	AJ235484	GQ983654	AF206919	AF479201
Grevillea robusta	AF197589.1			AF197577.1	
Grevillea baileyana		AF060434.1			
Grevillea banksii			AF542583.2		DQ008612.1
Halesia carolina	Z80190.1	DQ923988.1	DQ924097.1		AY727981.1
Halesia tetraptera				L49284.1	
Haloragis serra	U26325.2				
Haloragis aspera		AJ235488.2			
Haloragis glauca			EF179018		
Haloragis erecta				AF094547.1	AF274648.1
Hamamelis mollis	L01922.2				

Hamamelis helix		AF092105.1			
Hamamelis virginiana			AF013046.1	AF094551.1	AF036495.1
Harpagophytum grandidieri	HQ384883.1	HQ384722.1	HQ384517.1		
Hedycarya arborea	L12648.2	AJ235490.1	AM396509.1	AF206924.1	DQ008623.2
Hedyosmum arborescens	L12649.2	AJ235491.1	DQ401339.1	AF206925.1	AF479226.1
Heisteria parvifolia	AJ131771	AJ235492	DQ790199		DQ790232
Heisteria concinna				L24146	
Helianthus annuus	AF097517.1	AJ236205.1	DQ383815.1	AF107577.1	AF479183.1
Helwingia japonica	GQ436764	AF209596	AJ430195	U42524	
Helwingia sp.					GQ983586
Hernandia ovigera	L12650.2	DQ007419.1		DQ007407.1	
Hernandia nymphaeifolia			AJ247165.2		AY095462.1
Heteropyxis natalensis	U26326	AF209597	AF368208	AF206927	AM235609
Humiria balsaminifera	L01926.2	AJ235495.2	EF135549.1	AF206930.1	AY935815.1
Humulus lupulus	AF206777.1	AF209599.1	AY257528.1	AF206931.1	AF223066.1
Hydnocarpus heterophylla	AF206778	AF209600		AF206932	AF479121
Hydnocarpus sp.			EF135551		
Hydrastis canadensis	L75849.2	AF093382.1	AB069849.1	L75828.1	AF389268.1
Hydrocharis dubia	AB004892.1		AB002572.1	AY952398.1	
Hydrocharis morsus-ranae		DQ401334.1			
Hydrocotyle vulgaris	GQ983666	GQ983599	DQ133792		
Hydrocotyle sibthorpioides				X16605	
Hydrocotyle verticillata					AY189078
Hypericum perforatum	HQ332081.1	AF209602.1	DQ168438	AF206934.1	AF479122.1
Icacina mannii	AF206780	AF209603	HQ384577	AF206935.1	
Idesia polycarpa	AF206781	AF209604	AB233831.1	AB233623	AF479117.1
Idiospermum australiense	L12651.2	AJ235500.1	AY525342.1	AF206937.1	DQ008633.1
Ilex cornuta	GQ997347.1	GQ997300.1			
Ilex opaca			GQ248140.1	AF161010.1	AF479203.1
Ipomoea alba	AY100963.1				
Ipomoea purpurea		EU118126			
Ipomea batatas			AJ429355.1		
Ipomea hederacea				U38310.1	
Ipomea lacunosa					AF146016.1
Iris missouriensis	AY149365.1	AY147620.1			
Iris colchica			AY596632		
Iris tenax				JQ283937.1	JQ283879.1
Irvingia malayana	AF206782	AF209605	EF135553	AF206939	
Jasminum suavissimum	L01929.2			AJ236035	
Jasminum polyanthum		AJ235508.2			
Jasminum nudiflorum			AF531779.1		
Joinvillea plicata	EF423010	EF422979			
Joinvillea ascendens			AF164380	AF168855	
Juglans mandshurica	AY263932.1	AY263952.1			
Juglans nigra			AF118036	AF206943.1	AF479105.1
Kadsura japonica	AF197595.1	AF197607.1	DQ185525.1	AF293763.1	DQ008657.1
Kingdonia uniflora	AF093719.1	AF092115.1	FJ626519.1	AF094537.1	AF389245.1
Krameria ixine	EU644679	AJ235514	EU604050	AF206948	AF479116
Lacistema aggregatum	AF206787.1	AF209613.1	AY935933.1	AF206949.1	AY935816.1
Lactoris fernandeziana	L08763.1	AJ235515.2	DQ882195.1	U42783.1	AY095463.1

Lamium amplexicaule	AB266225	AJ236165		L49287	
Lamium album			AJ429332		
Larrea tridentata	AF200474	AY935860	AY935935	AY929372	AY935818
Leea guineensis	AJ235783.1	AJ235520.2	AF274621.1	AF206951.1	AF274653.1
Licania elaeosperma	AB233846	AB233638	AB233742	AB233534	
Licania heteromorpha					AF222370
Lilium superbum	L12682.2	AF209618.1		AF206952.1	
Lilium alexandrae			AB030849.1		
Lilium michauxii					AF205126.1
Linnaea borealis	AF446941	GQ983619	AF446911		
Linum arboreum	AY380351				
Linum perenne		AJ235521	AB038182	L24401	
Linum usitatissimum					EU307117.1
Liquidambar styraciflua	AF119181.1	AF092104.1	AF133219.1	U42553.1	AF479217.1
Lissocarpa benthamii	EU980793.1	DQ923969.1	DQ924077.1		AY727956.1
Lonicera japonica	EU201192	GQ983602			GQ983587
Lonicera orientalis			AJ430196.1		
Lonicera maackii				U66701.1	
Lotus corniculatus	U74213		HM049505	BV165278.1	
Lotus japonicus		AP002983			
Luculia gratissima	AM117243.1	HQ384771	AJ429325		
Lythrum hyssopyfolia	L10218.1		HM850986.1		
Lythrum salicaria		AF209621.1		AF206955.1	AF479240.1
Magnolia tripetala	AF206791.1	AJ235526.1		AF206956.1	
Magnolia denudata			AF123465.1		AF479244.1
Mahonia bealei	L12657.2	AF197611.1		AF293755.1	DQ008613.1
Mahonia japonica			AF542585.1		
Malesherbia linearifolia	AF206792	AF209622	EF135562	AF206957	DQ123011
Malpighia emarginata	AF344493.1		AF344561.1		
Malpighia coccigera		AJ235527.2		L24046.1	
Manilkara zapota	AF213793.1	AJ235528.1	DQ924092.1	L49288.1	AY727946.1
Marcgravia rectiflora	Z83148.1	AJ235529.1			AY727937.1
Marcgravia sp.			AJ429289.1		
Mayaca fluviatilis	AF036885.1	AF168929.1			AF293855.1
Mayaca aubletti				AF168859.1	
Maytenus senegalensis	AY380353	AY788272			
Maytenus floribunda			EU328961		EU328667.1
Maytenus arbutifolia				AY674616	
Mazus stachydifolius	EU348860.1				
Mazus reptans		HQ384705.1			
Mazus rugosus			FN773547.1		
Mazus pumilus				GU359050.1	AF148278.1
Medusagyne oppositifolia	Z75670	AJ235530	FJ670030	AF206959	AF479120
Melanophylla alnifolia	U50254.1	AF209625.1		AF206960.1	AY189085.1
Melanophylla sp.			AJ429373.1		
Meliosma veitchiorum	AF206793.1	AF209626.1	AJ581449.1	AF206961.1	AF389271.1
Menispermum dauricum	FJ026493.1				
Menispermum canadense		AF093384.1	GU266604.1	L75834.1	AF389257.1
Menyanthes trifoliata	L14006.2	AJ235533.2	AJ429386.1	AJ236009.1	AF479185.1
Mesua sp.	AF206794	AF209627	EF135567	AF206962	
Metanarthecium luteoviride	AB088837.1	AF308041.1	AB040163.1	AF309410.1	

Metasequoia glyptostroboides	AJ235805.1	AJ235534.2	AB030122	EF053174.1	DQ008663.1
Microdesmis pierlotiana	AY663645			AY674618	
Microdesmis puberula		AY788238	EF135570		
Mimosa speggazzinii	Z70151			X66686	X66762
Mimosa tenuiflora		EU811854			
Mimosa pudica			AY177668		
Misodendrum linearifolium	L26074.1		DQ787438.1	L24397.2	DQ790211.1
Mitchella repens	AF190440	AF209630	HQ593366.1	U42802	AF148279
Mollugo verticillata	M62566.1	AF209631.1	AY936330.1	U42828.1	AF479088.1
Morella cerifera	AJ626759.1			AF206967.1	AF479247.1
Morella pensylvanica			AY491657.1		
Moschopsis rosulata	X87390	AJ318979	GQ983662		
Moultonianthus leembruggianus	AY794982	FJ669974.1	FJ670015.1	FJ669695.1	
Myodocarpus involucratus	AY188430				AY189095
Myodocarpus fraxinifolius		GQ983630		GQ983570	
Myodocarpus simplicifolius			AF271754		
Myrtus communis	AF294254.2	JF268426	AM490009.1		EU002154.1
Nelsonia sp.	HQ384879.1	HQ384717.1	HQ384513.1		
Nelumbo lutea	DQ182337.1	AF528853.1		AF094556.1	AF389259.1
Nelumbo nucifera			AF543740.1		
Nepenthes alata	L01936.2	AJ235542.2			
Nepenthes madagascariensis			AF315883.1		
Nepenthes sp.				U42787.1	AF389260.1
Nerium oleander	AJ002886.1	AJ236189.1	Z98173.1	AF107572.1	AF479178.1
Neurada procumbens	U06814	AF209637		AF206970	AF479225
Nicotiana tabacum	Z00044.2	AF035909.1		AJ236016.1	AF479172.1
Nicotiana acaulis			AB039985.1		
Nothofagus antarctica	AY263939	AY147106		AY147111	
Nothofagus solandri			AB015464.1		
Nothoscordum bivalve	Z69202.1				AF293853.1
Nothoscordum dialystemon		AJ235504.2	JQ435522.1		
Nymphaea odorata	M77034.1	AJ235544.1	DQ185549.1	AF206973.1	
Nymphaea sp.					AY095465.1
Nymphoides peltata	EF173110		GQ983659		
Nymphoides geminata		AJ236204		AJ236010.1	AF479186.1
Nyssa ogeche	L11228.2		U96886.1	U52032.1	AF297545.1
Nyssa sylvatica		AJ235545.2			
Ochthocosmus longipedicellatus	FJ707535.1	FJ707529.1	EF135573	AY674621	
Olea europaea	DQ673304.1	AJ236163.1	AJ429335.1	L49289.1	AF479171.1
Opuntia dillenii	AY875233				
Opuntia microdasys		HQ843258		HQ843432	HQ843456
Opuntia quimilo			AY015279		
Panax quinquefolius	U50250.1	AJ236210.1			AF479193.1
Panax ginseng			AB044903.2		
Panax zingiberensis				AB085764.1	
Paracryphia alticola	AJ402983	AJ419679	AJ429367	GQ983571	GQ983589
Patrinia triloba	AF446951	GQ983625	AF446921	GQ983572	
Paulownia tomentosa	L36447.1	AJ236174.1	AJ429339.1	AJ236039.1	JF321129.1

Penthorum sedoides	L11197.2	AJ235555.2	EF179063.1	U42825.1	AF274661.1
Perrottetia ovata	AY935737	AY935842	AY935916	AY929358	AY935806
Petalonyx nitidus	AF299086.1	AJ236232.1		AJ235989.1	AY260028.1
Petalonyx crenatus			AF503295.1		
Phlox longifolia	AF206809.1	AJ236221.1			
Phlox gracilis			L34203.1		
Phlox hoodii				L49293.1	
Phlox divaricada					AF148281.1
Photinia fraseri	L11200.2	AF209653.1		U42800.1	AF479101.1
Photinia serrulata			AF288111.1		
Physena sp.	Y13116				
Physena madagascariensis		HQ843260		HQ843434	HQ843458
Pimelodendron zoanthogyne	AJ418812	AY788247	EF135582	AY674628	
Pinus krempfii	DQ353718.1				
Pinus thunbergii		NC_01631.1			
Pinus wallichiana			EF546729.1		
Pinus taeda				AH001728.1	
Pinus strobus					AY056501.1
Plantago lanceolata	L36454.1	AF209656.1		AJ236046.1	
Plantago argentea			AJ429344.1		
Plantago virginica					AF148277.1
Platanus occidentalis	AF081073.1	AF528858.1	AF543747.1	U42794.1	AF274662.1
Plumbago auriculata	M77701.1		EU002187	U42795.1	AF036492.1
Plumbago zeylanica		AJ235565.2			
Polygala cruciata	L01945.2	AJ235568.1			AF479233.1
Polygala californica			AY386842.1		
Polygala pauciflora				U42797.1	
Pontederia cordata	U41593.1	AF209657.1	AF434872.1	AF206998.1	AF205517.1
Populus tremuloides	M58392.1	AF209658.1		AF206999.1	AF479118.1
Populus nigra			AB038186.1		
Proboscidea louisianica	L01946.2	AJ236161.1	AF531809.1	AJ236040.1	
Pseudonemaclads oppositifolius	AJ318992	EU437600	GQ983646		
Puya raimondii	AF206814	AF209661	FJ968207	AF207001	
Qualea sp.	U02730.1	AF209662.1		AF207002.1	
Qualea grandiflora			AF368216.1		
Quillaja saponaria	U06822	GQ497659.1	AY386843		
Quintinia verdonii	AF299092		AJ429366		
Quintinia quatrefagesii		AJ318983		GQ983576	GQ983590
Ranunculus acris	AY395557.1		AY954199.1		
Ranunculus macranthus		DQ069346.1			
Ranunculus sceleratus				EF526357.1	
Ranunculus keniensis					AF389269.1
Rehmannia elata	HQ384874.1	HQ384709.1	HQ384505.1		
Restio tetraphyllus	AF206816.1		AF164379.1	AF207006.1	AF486829.1
Restio paludosus		AJ419152.1			
Rhamnus lycioides	AJ390070.1				
Rhamnus catharticus		AJ235579.2			
Rhamnus cathartica			AY257533.1	AJ235979.1	
Rhamnus utilis					JF317393.1
Rhizophora stylosa	AF127686		AF105092	AY289627	
Rhizophora racemosa					AF224693

Rhododendron hippophaeoides	L01949.2			AF419807.1	
Rhododendron impeditum		AY725932.1			AY727973.1
Rhododendron wadanum			AB012746.1		
Rhus copallina	U00440	AF035912			
Rhus transvaalensis			EU214283		
Rhus typhina				GU476470	
Rivina humilis	M62569	HQ843261	AY042646	HQ843438	HQ843461
Roridula gorgonias	L01950.2	AJ236180.1	AJ429294.1	AF207010.1	AY727965.1
Roussea simplex	AF084477.1	AJ235586.2	AJ429389.1	U42548.1	AF479243.1
Sabia swinhoei	FJ626616.1	AF093395.1		L75840.1	AF389272.1
Sabia japonica			AM396512.1		
Salix reticulata	AJ235793	AJ235590	EF135592	AF207011	
Sambucus racemosa	AF446928.1		AF446898.1		
Sambucus caerulea		GQ983634			
Sambucus canadensis				AF193941.1	
Sarcandra chloranthoides	AY236833.1		AJ966796.1		DQ008655.1
Sarcandra grandiflora		AJ235593.1		AF207012.1	
Sarcobatus vermiculatus	AF132088	HQ843262	AY042652	GQ497586.1	HQ843462
Sarracenia flava	L01952.2	AJ235594.2			
Sarracenia purpurea			AJ429296.1	U42804.1	AY260044.1
Saururus cernuus	L14294.1	AF187061.1	AF543749.1		AY095468.1
Saururus chinensis				AY572303.1	
Sauvagesia africana	AB233909				
Sauvagesia erecta		FJ669984	EF135593	EF135604	EF135618
Scaevola aemula	EU017199.1	EU017162.1	EU385394	AJ236008.1	
Scaevola auriculata					GQ983591.1
Schotia afra	AM235016				
Schotia brachypetala			EU362038	X66779	X66764
Sesamum indicum	HQ384882	HQ384721	AJ429340	AJ236041	
Siphonodon celastrineus	AF206821.1	AF209676.1	AY935919.1	AF207021.1	
Siphonodon australis					AF222346.1
Sloanea latifolia	AF022131.1			U42826.1	
Sloanea berteriana		AJ235603.2			AF479126.1
Sloanea australis			AY935938.1		
Smilax glauca	AF206822.1	AF209677.1		AH003500.1	
Smilax china			AB040204.1		
Smilax bonanox					AF293852.1
Sparganium americanum	M91633.1				
Sparganium eurycarpum		AY465539.1		L24419.1	
Sparganium glomeratum			AY952426.1		
Spathiphyllum wallisii	AJ235807.1	AJ235606.2	AM920559.1	AF207023.1	AY095473.1
Spiraea vanhouttei	L11206.2			U42801.1	
Spiraea betulifolia		AJ235608.2			AF479103.1
Spiraea cantoniensis			AF288127.1		
Staphylea trifolia	AY646111.1	AJ235611.2		AJ235978.1	AF479133.1
Staphylea colchica			EU002189.1,		
Stellaria media	AF206823.1	AF209680.1	AY936299	AF207027.1	AF479084.1
Sterculia lanceolata	AY082362.1				
Sterculia apetala		AJ233089.1			AF479137.1

Sterculia tragacantha			AY321178.1		
Sterculia recordiana				AF207029.1	
Stylidium graminifolium	L18790	AJ236201		AJ236011	AJ236011
Stylidium majus			AF542576		
Stylobasium spathulatum	U06828	AF209681	EU604032	AF207030	
Swietenia macrophylla	U39080.2	AJ235616.2	AY128200.1	AF207031.1	AF479241.1
Symphonia tanalensis	AB233852.1	AB233644.1	AB233748.1	AB233540.1	
Symphoricarpos albus	L11682			U42513.1	AF479242.1
Symphoricarpos sp.		GQ983633			
Symphoricarpos occidentalis			GU168654		
Tapiscia sinensis	AF206825	AF209685	EU002190	AF207034	AF479146
Tapura fischeri	GQ424471				
Tapura guianensis		FJ669972.1	FJ670009.1	FJ669693.1	
Tasmannia insipida	L01957.2	AF093424.1		AF207035.1	AY095469.1
Tasmannia lanceolata			DQ882241.1		
Terminalia catappa	U26338	AF209686	GU135057	AF207037	
Terminalia boivinii					AF479147
Ternstroemia stahlii	AF206827.1	AF209687.1		AF207038.1	AY727955.1
Ternstroemia longipes			AF380110.1		
Tetracera asiatica	AJ235796	AJ235622	AY042665	AJ235982	AF479097
Tetrameles nudiflora	L21943.1	AF209689.1	AY968458.1	U41502.1	AY968422.1
Tetrapanax papyriferus	U50256	GQ983624	GQ434267		AY189111.1
Tetrapterys glabrifolia	AB233903	AB233695	AB233799	AB233591	
Thunbergia usambarica	L12596.1				
Thunbergia coccinea		AJ235625.2			
Thunbergia alata			AF531811.1	AF107569.1	
Thymelaea hirsuta	Y15151	AJ235626	EU002191	AF207041	AF479234
Toricellia tiliifolia	AF299089		AY188410		AY188410
Tradescantia pallida	AM110254.1		AM114727.1		
Tradescantia ohiensis		AF168950.1		AF069213.1	
Tragopogon pratensis	AY395563.1				
Tragopogon dubius		AJ236197.1	AJ633258.1	U42502.1	AF036493.1
Trichilia emetica	U39082.2	AJ235629.2	AY128202.1	AF207045.1	AY128171.1
Trigonia nivea	AF206830		EF135598	AF207047	
Trigonia boliviana		AB233640			
Trimenia moorei	AF121367.2	AY116653.1	DQ401360.1		AY095470.1
Tripterygium regelii	AY788193	AY788260	EU328941.1	AY788161	AF222351.1
Trochodendron aralioides	L01958	AF093423	GQ998807	AH001800	AF479205
Turnera subulata	DQ123398				DQ123012
Turnera ulmifolia		AJ235634	EF135599	U42817	
Vahlia capensis	L11208.2	AJ236217.1	AJ429316.1	U42813.1	AF479182.1
Valeriana officinalis	L13934.1	AJ235637.2		AJ236003.1	AF479199.1
Valeriana hirtella			AJ429396.1		
Viburnum acerifolia	AF446927.1			AJ236007.1	
Viburnum opulus		AJ235640.2			
Viburnum rhytidophyllum			AJ429391.1		
Viburnum erosum					JF321130.1
Viola philippica	AB354436		DQ842600	AB354544.1	
Viola pubescens		FJ669992			EF564797
Vitis aestivalis	L01960.2				AF479207.1
Vitis vinifera		DQ424856.1	DQ424856.1		
Vitis sp.				AF207053.1	
Weigela hortensis	AF446938	GQ983609	AF446908		

Welwitschia mirabilis	AJ235814.1	NC_010654.1	AF280996.1	AH001808.1	DQ008662.2
Xanthosoma sagittifolium	L10246.2				
Xanthosoma mafaffa		DQ401330.1			
Xanthosoma helleborifolium			AM920612.1		
Xerophyta retinervis	EU213532.1	KT204786.1	EU214302.1		AF205878.1
Xerophyta humilis				EF418586.1	
Zea mays	Z11973.1	NC_001666.2	NC_001666.2	AH001709.1	
Zelkova serrata	AF206835	AF209699.1		U42819.1	AF479099.1
Zelkova schneideriana			AF345328.1		

APPENDIX 2

Table A2.1. Table of sampled fossils in Osmundaceae.

Fossil		Clade				
Taxon	**Age**	**Osmundaceae**	***Osmunda***	**JV**	**JR**	**BW**
Osmunda cinnamomea	72-83	✓	✓			
Osmunda cinnamomea	12-14	✓	✓			
Osmunda cinnamomea	16	✓	✓			
Osmunda precinnamomea	56-66	✓	✓			
Osmunda chengii	145-160	✓				
Osmunda bromeliaefolia	11-14	✓	✓	✓		
Osmunda arnoldii	56-66	✓	✓	✓		
Osmunda dowkeri	34-40	✓	✓	✓		
Osmunda dowkeri	56-59	✓	✓	✓		
Osmunda ilianensis	4-6	✓	✓	✓		
Osmunda nathorstii	30-40	✓	✓			
Osmunda oregonensis	34-40	✓	✓			
Osmunda pluma	56-66	✓	✓			
Osmunda shimokawaensis	12-14	✓	✓	✓	✓	
Osmunda wehrii	16	✓	✓			
Osmunda pulchella	183-185	✓				
Osmunda liaoningensis	153-165	✓	✓			
Osmunda wangii	153-165	✓	✓			
Osmunda plumites	153-165	✓				
Osmunda clatoniites	145-160	✓				
Osmunda delawarensis	84-86	✓	✓	✓		
Osmunda diamensis	157-166	✓				
Osmunda japonica fossilis	34-40	✓	✓	✓		
Osmunda lignita	44-49	✓	✓	✓		
Osmunda lignita	28-34	✓	✓	✓		
Osmunda macrophylla	56-59	✓	✓	✓		
*Osmunda macrophylla**	56-59	✓	✓	✓		
Osmunda parschlugiana	11-13	✓	✓	✓		
Osmunda strozzii	5-12	✓	✓	✓		
Osmunda palaeobanksiifolia	28-34	✓	✓	✓		
Osmunda macrophylla	28-41	✓	✓	✓		
Osmunda vancouverensis	129-140	✓	✓			
Osmundopsis sturii	169-170	✓				
Todea amissa	52	✓				✓
Osmundopsis plectrophora	190-201	✓				

* multiple specimens from same sediment (see Grimm et al. 2015)

Table A2.2. Table of sampled fossils in CS

Fossil		Clade	
Taxon	**Age**	**Solanaceae (inc. stem branch)**	**Solanoideae (inc. stem branch)**
Datura cf. stramonium	2.6-3.6	✓	✓
Hyoscyamus sp.	2.6-5.3	✓	✓
Hyoscyamus niger	2.6-3.6	✓	✓
Physalis alkekengi	3.6-7.3	✓	✓
Physalis alkekengi	2.6-3.6	✓	✓
Physalis alkekengi	2.6-3.6	✓	✓
Physalis pliocenica	23.0-28.4	✓	✓
Physalis pliocenica	2.6-5.3	✓	✓
Physalis pliocenica	7.3-11.6	✓	✓
Physalis pliocenica	11.6-13.7	✓	✓
Physalis pliocenica	11.6-13.7	✓	✓
Physalis alkekengi	3.6-7.3	✓	✓
Physalis aff. alkekengi	3.6-5.3	✓	✓
Scopolia carniolica	2.6-3.6	✓	✓
Solanispermum reniforme	46.0-48.0	✓	
Solanispermum reniforme	46.0-48.0	✓	
Solanispermum reniforme	40.0-44.0	✓	
Solanispermum reniforme	37.2-40.4	✓	
Solanispermum reniforme	28.4-33.9	✓	
Solanum arnense	46.0-48.0	✓	
Solanum cf. persicum	2.6-3.6	✓	✓
Solanum dulcamara	3.6-7.3	✓	✓
Solanum dulcamara	2.6-3.6	✓	✓
Solanum dulcamara	2.6-3.6	✓	✓
Solanum dulcamara	2.6-3.6	✓	✓
Solanum dulcamara	2.6-5.3	✓	✓
Solanum dulcamara	2.6-3.6	✓	✓
Solanum dulcamara	2.6-3.6	✓	✓
Solanum dulcamara	2.6-5.3	✓	✓
Solanum nigrum	5.3-11.6	✓	✓
Solanum sp.	2.6-5.3	✓	✓

Table A2.3. Table of sampled fossils in Spermatophyta. All fossils listed here are node-calibration-fossils. CG refers to the fossils being used to calibrate the Crown Group of the listed clade. SG refers to the fossil being used to calibrate the Stem Group of the listed clade.

Fossil		
Taxon	**Age**	**Clade**
Cycas sp	323.2-358.9	CG Spermatophyta
Several	129.4-136.0	CG Angiospermae
Monitianthus mirus	108.0-113.0	SG Nymphaceae
Anacostia marylandensis	108.0-113.0	SG Schisandraceae
Asteropollis sp.	121.0-125.0	SG Hedyosmum
Sarcandra	99.6-103.7	SG Sarcandra
Walkeripollis gabonensis	125-127.2	SG Winteraceae

Lactoripollenites africanus	72.1-93.6	SG Lactoris
Saurus tuckerae	37.8-47.8	SG Saurus
Endressinia brasiliana	108.0-113.0	CG Magnoniales
Futabanthus asamigawaensis	88.05-89.8	SG Annonaceae
Virginianthys calycanthoides	104.7-113.0	SG Calycanthacaceae
Potomocanthus lobatus	104.7-113.0	SG Lauraceae
Liliacidites sp.	113.0-125.0	CG Monocotyledoneae
Mayoa portugallica	121.0-125.0	SG Mondteroideae
Cardstonia toldmanii	69.05-77.85	CG Alismataceae
Cyclanthus messelensis	37.8-47.8	SG Cyclanthaceae
Smilacoideae	41.2-47.8	SG Smilaceae
Sabal magothiensis	83.6-86.3	SG Arecaceae
Monochoria	41.2-47.8	SG Pontederiaceae
Spirematospermum chandlerae	77.9-83.6	SG Zingiberales
Typha protogaea	66.0-72.1	SG Typhaceae
Restio subverticillatus	66.0-72.1	SG Restionaceae
Monoporites annulatus	56.0-66.0	SG Poaceae
Several	125.0-129.4	SG Eudicotyledoneae
Teixeiraea lusitsnica	112.0-118.0	CG Ranunculales
Sargentodoxa globosa	37.8-47.8	SG Larzidabalaceae
Prototinomiscium testudinarum	66.0-72.1	SG Menispermaceae
Menispina evidens	56.0-61.6	SG Cissampelos
Mahonia sp.	33.9-56.0	CG Beberidaceae
Paleoactaea nagelii	56.0-59.2	SG Ranunculaceae excluding Glaucidium and Hydrastis
Meliosma prealba	66.0-71.2	CG Sabiaceae
Nelumbites extenuinervis	100.5-106.75	SG Nelumbonaceae
Sapindopsis variablis	104.7-108.9	SG Platanaceae
--	100.5-104.7	CG Pentapetalae
Hibbertia sp.	47.8-56.0	CG Dilleniales
Triprojectacites-Aquilapollenites complex	69.05-77.85	SG Santalales
Gothanipollis sp.	66.0-67.0	SG Loranthaceae
Polygonocarpum johnsonii	66.0-72.1	SG Polygonaceae
Caryophylloflora paleogenica	33.9-47.8	CG Caryophyllaceae
Coahuilacarpon phytolaccoides	72.1-75.9	CG Phytolaccaceae
Hironoia fusiformis	88.05-89.8	SG Cornaceae plus Nyssaceae
Beckettia pyriformis	66.0-72.1	SG Nyssaceae
Tylerianthus crossmanensis	89.8-93.9	SG Hydrangaceae
Pentapetalum trifasciculandricus	89.8-93.9	CG Ericales
Gilisenium hueberi	37.8-47.8	CG Polemoniaceae
Rehderodendron stonei	33.9-37.8	SG Styracaceae
Paleoekianthus sayrevillensis	33.9-37.8	SG Clethraceae, Cyrillaceae, and Ericaceae
Parasaurauia allonensis	83.6-84.95	CG Actinidiaceae
Icacinoxylon alternipunctata	89.8-93.9	SG Icacinaceae
Eucommia eocenica	37.8-47.8	SG Eucommiaceae
Scandianthus costatus	79.8-84.95	SG Vahliaceae
Solanispermum reniforme	33.9-56.0	SG Solanaceae
Emmenopterys dilcheri	37.8-47.8	CG Rubiaceae

Apocynospermum coloradensis	37.8-47.8	SG Apocynaceae
Pistillipollenites macgregorii	33.9-56.0	CG Gentianaceae
Fraxinus excelsior	5.33-11.63	CG Oleaceae
Gratiola tertiaria	5.33-23.03	CG Plantaginaceae
Trapella cf. antennifera	2.58-5.33	CG Pedaliaceae
Catalpa sp.	38.1-39.5	CG Bignoniaceae
Acanthus rugatus	27.82-30.86	CG Acanthaceae
Paulownia inopinata	11.63-15.97	SG Paulowiniaceae
Ajuginucula smithii	27.82-30.86	CG Lamiaceae
Ilex hercynica	61.6-66.0	SG Aquifoliaceae
Campanula palaeopyramidalis	5.33-23.03	CG Campanulaceae
Menyanthes cf. trifoliata	5.33-23.03	CG Menyanthaceae
Raiguenrayun cura	47.6	SG Asteraceae
Psilatricolporites protrudens	9.7-10.3	CG Calyceraceae
Toricellia bonesii	56.0-59.2	CG Torricelliaceae
Paleopanax oregonensis	37.8-47.8	CG Araliaceae
Diervilla echinate	27.82-33.9	CG Diervillaceae
Patrinia paleosibirica	5.33-23.03	CG Valerianaceae
Divisestylus brevistamineus	89.8-93.9	SG Itaceae
Tarahumara sophiae	72.1-83.6	SG Haloragaceae
Ampelocissus parvisemina	56.0-66.0	CG Vitaceae
Esgueiria futabensis	88.05-89.8	SG Combretaceae
Decondon tiffneyi	72.1-75.9	CG Lythraceae
--	83.6-84.95	SG Myrtaceae
Paleomyrtinaea princetonensis	56.0-59.2	CG Myrtaceae
Turpinia uliginosa	33.9-41.2	CG Staphyleaceae
Bursericarpum aldwickense	47.8-56.0	CG Burseraceae
Aesculus hickeyi	56.0-66.0	CG Sapindaceae
Toona sulcata	47.8-56.0	CG Meliaceae
Tapiscia occidentalis	37.8-47.8	CG Tapisciaceae
Florissantia ashwilii	37.8-47.8	CG Malvales
Malviciphyllum macondicus	56.0-61.6	SG Malvaceae
Dressiantha bicarpelata	89.8-93.9	SG Brassicales
Thlaspi primaevum	23.0-27.8	CG Brassicaceae
Zygophyllocarpum sp.	23.0-27.82	SG Zygophyllaceae
Leguminocarpon gardneri	56.0-66.0	SG Fabaceae
--	56.0-59.2	SG Papilionoideae
Prunus wutuensis	47.8-56.0	CG Rosaceae
Coahuilanthus belindae	72.1-77.85	SG Rhamnaceae
Nothofagidites senectus	82.8-84.2	SG Nothofagaceae
Normapolles complex	97.2	SG Core Fagales.
Castaneoidea puryearensis	37.8-47.8	CG Fagaceae
Caryanthus sp.	83.6-84.95	SG Juglandaceae plus Myricaceae
Polyptera manningi	64.4	CG Juglandaceae
Bedellia pusilla	83.6-84.95	SG Betulaceae
Cucurbitospermum sheppeyense	47.8-56.0	SG Cucurbitaceae
Celastrus sp.	37.8-47.8	CG Celastraceae
Tripterygium cf. regelii	2.58-5.33	SG Tripterygium
Platydiscus peltatus	79.8-84.95	SG Cunoniaceae
Eucryphia falcata	56.0-59.2	CG Cunoniaceae
Sloanea ungeri	61.6-66.0	CG Elaeocarpaceae
Paleoclusia chevalieri	89.8-93.9	SG Clusiaceae
Populus tidwelli	37.8-47.8	SG Salix plus Polulus

Populus wilmattae	37.8-47.8	SG Populus
Eoglandulosa warmanensis	37.8-47.8	SG Malpighiaceae
Tetrapterys harpyiarum	27.82-33.9	SG Tetrapterys
Hippomaneoidea warmanensis	37.8-47.8	SG Euphorpioideae

APPENDIX 3

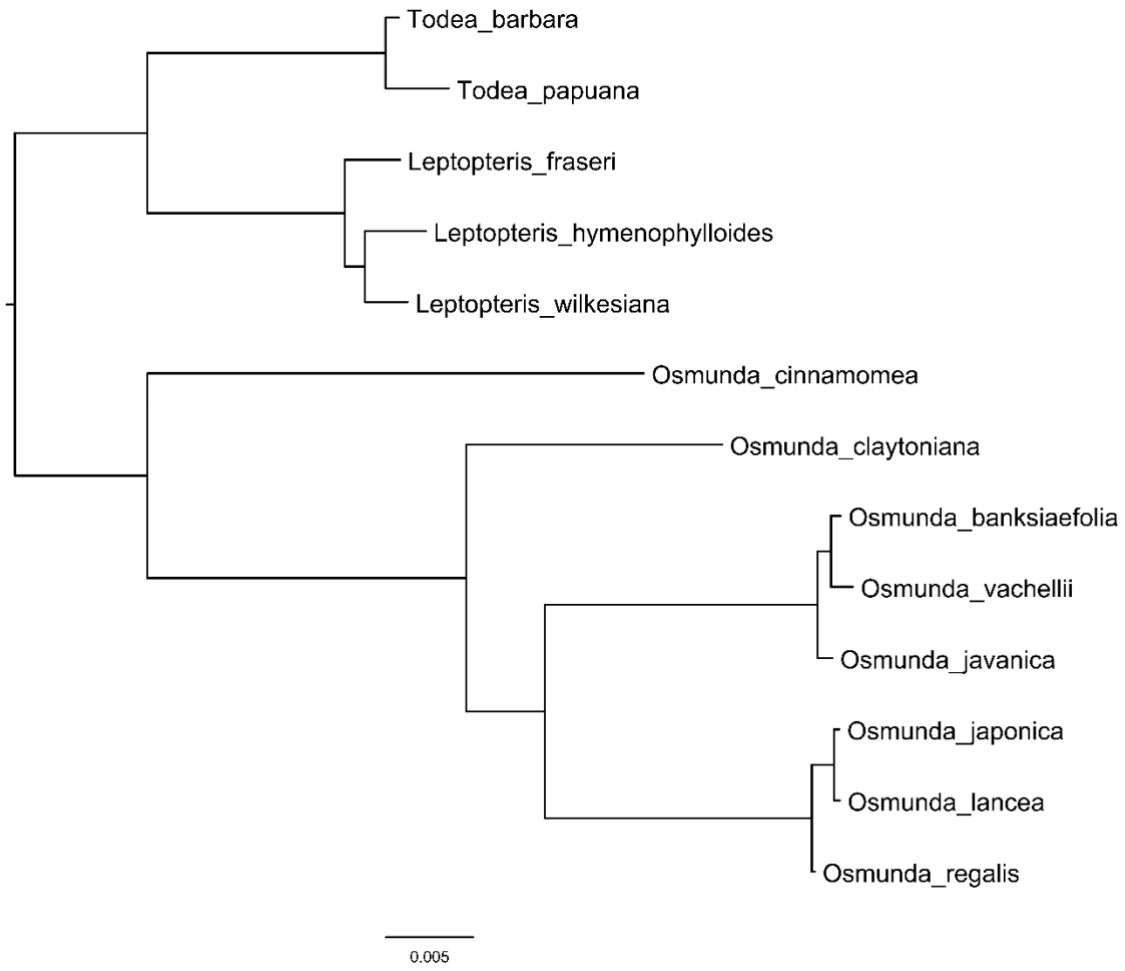

Figure A3.1. Inferred phylogeny for Osmundaceae. Ultrametric phylogeny and subsequent divergence time analyses are constrained to this topology.

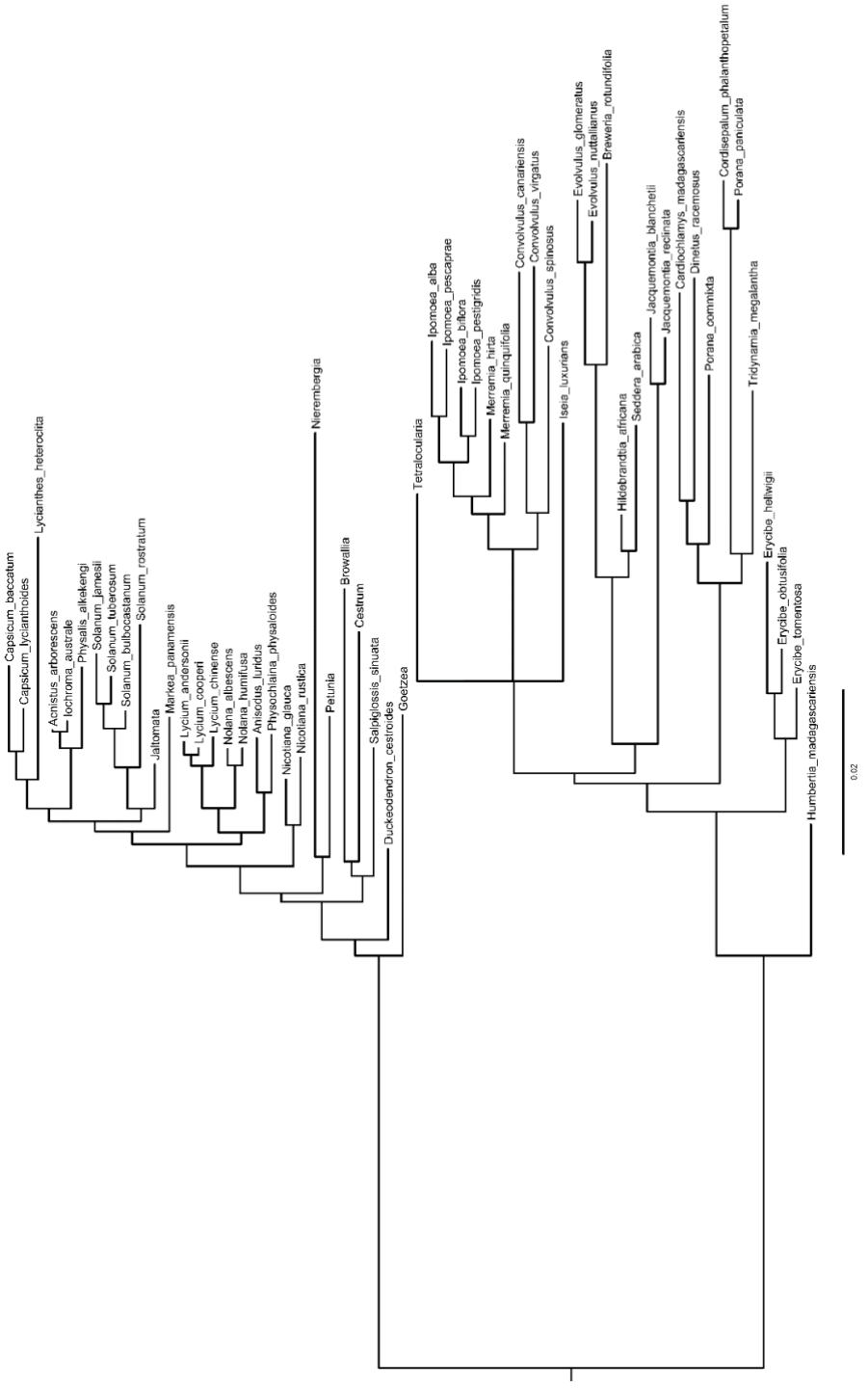

Figure A3.2. Inferred phylogeny for CS. Ultrametric phylogeny and subsequent divergence time analyses are constrained to this topology.

Continued on following page

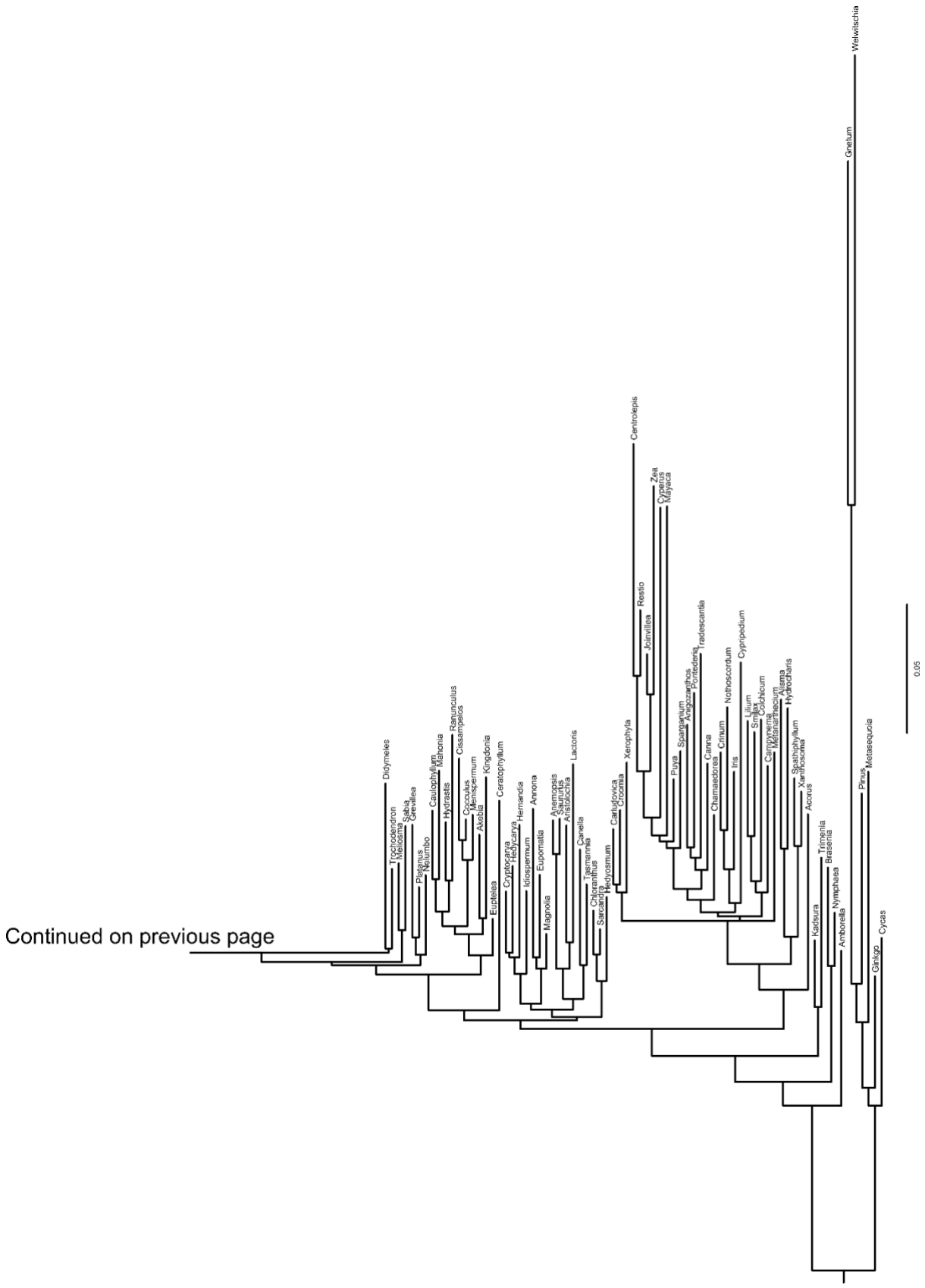

Figure A3.3. Inferred phylogeny for Spermatophyta. Ultrametric phylogeny is constrained to this topology.

APPENDIX 4

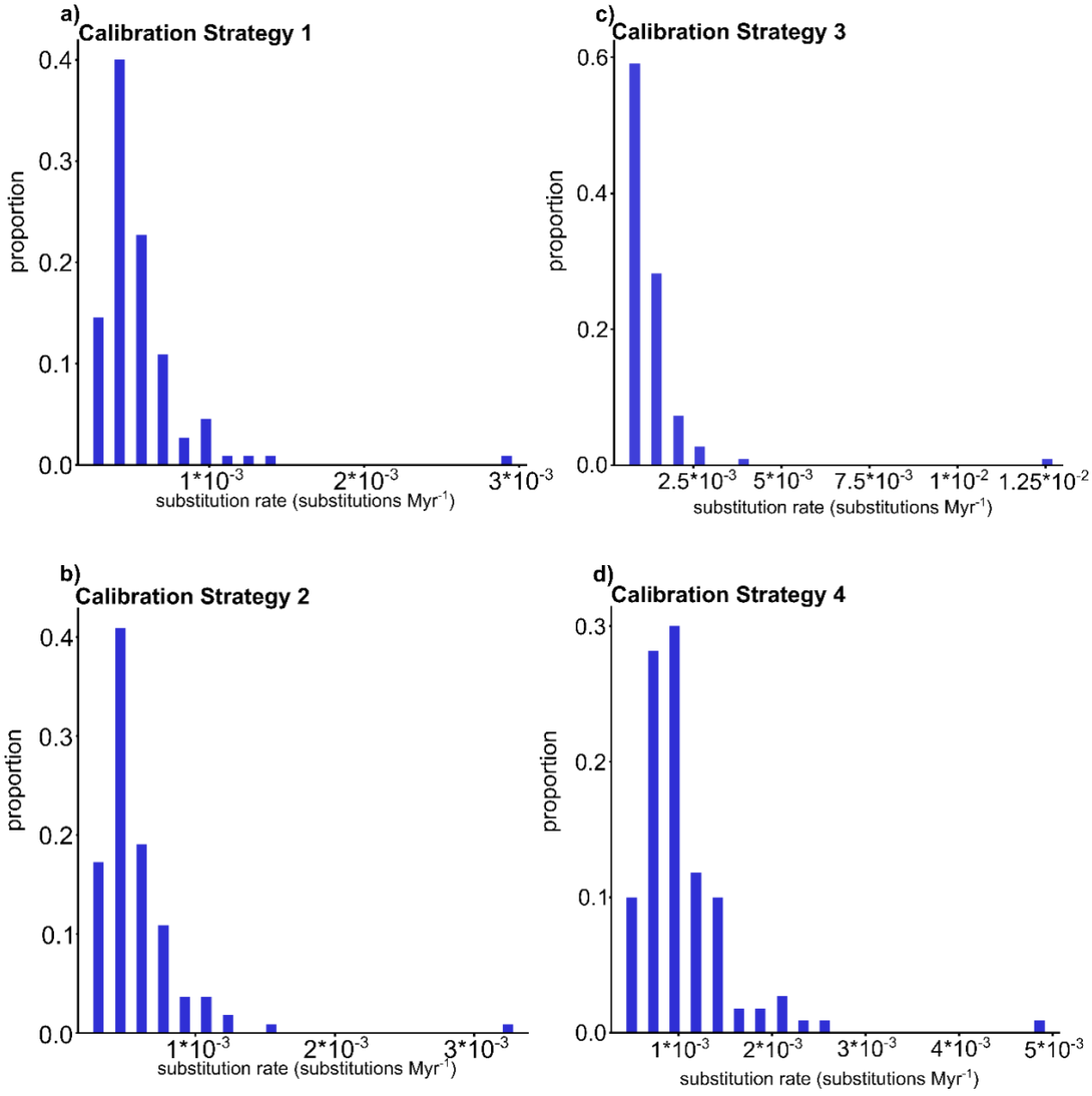

Figure A4.1. The distribution of inferred branch specific substitution rates in Convolvulaceae and Solanaceae when different calibration strategies are used. The proportion refers to the proportion of branches with substitution rates within each rate bin. Each bin is equal to 0.05 of the range between the minimum and maximum rate. a) – d) show the distribution for Calibration Strategies 1-4 respectively.

APPENDIX 5

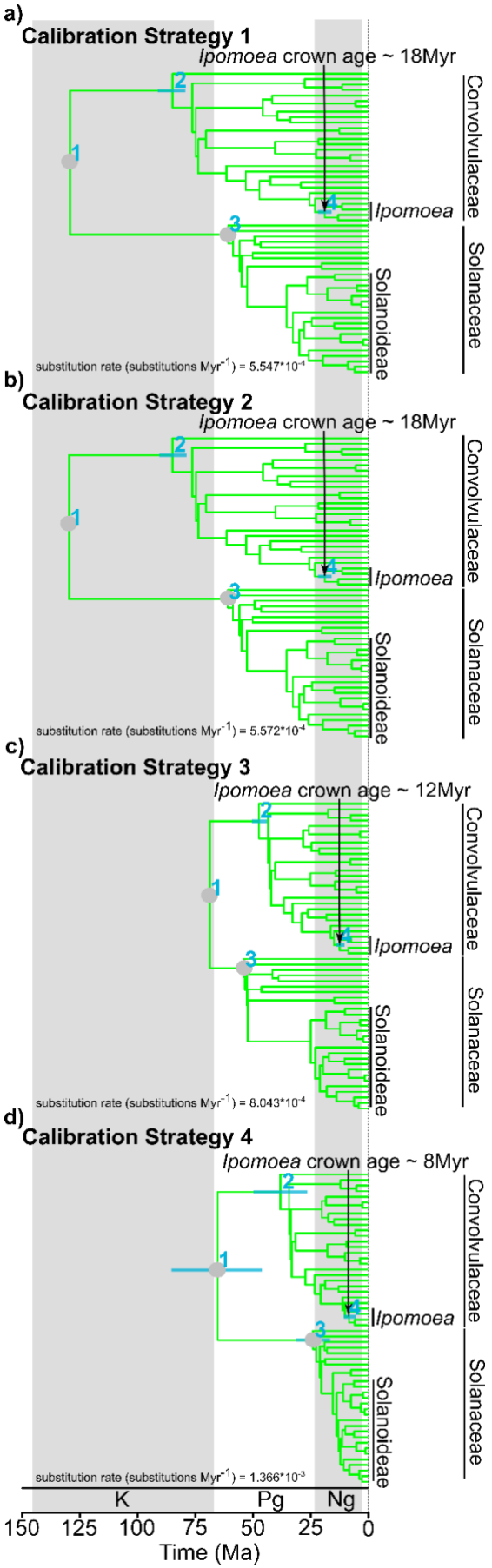

Figure A5.1. Time-calibrated phylogenies for Convolvulaceae and Solanaceae inferred with different calibration strategies and a strict clock. Grey circles refer to calibrated nodes as set out in Figure 4.1. Blue bars refer to 95% HPDs for key clades as set out in Figure 4.1. a) – d) refer to Calibration Strategies 1-4 respectively. These are implemented as set out in Figure 4.1.

APPENDIX 6

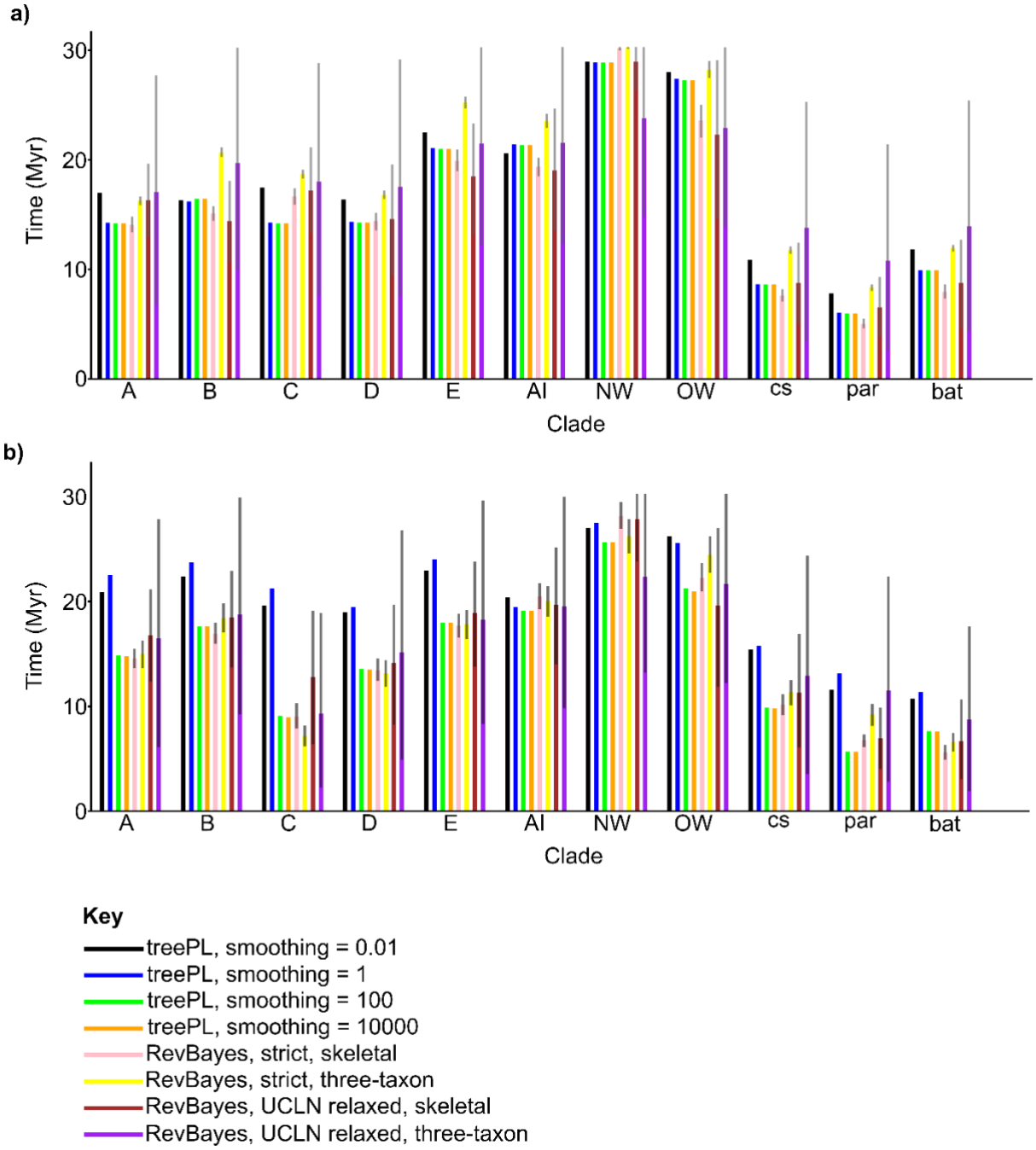

Figure A6.1. Inferred ages for major clades within *Ipomoea* in nuclear a) and chloroplast b) time-calibrated phylogenies. Major clades are labelled as set out in Figure 4.2. Different coloured bars refer to age estimates derived from different inference methods (see key). For treePL these age estimates refer to point age estimates, for RevBayes these age estimates refer to MPEs. Where RevBayes has been used, the grey bar refers to the 95% HPD.

APPENDIX 7

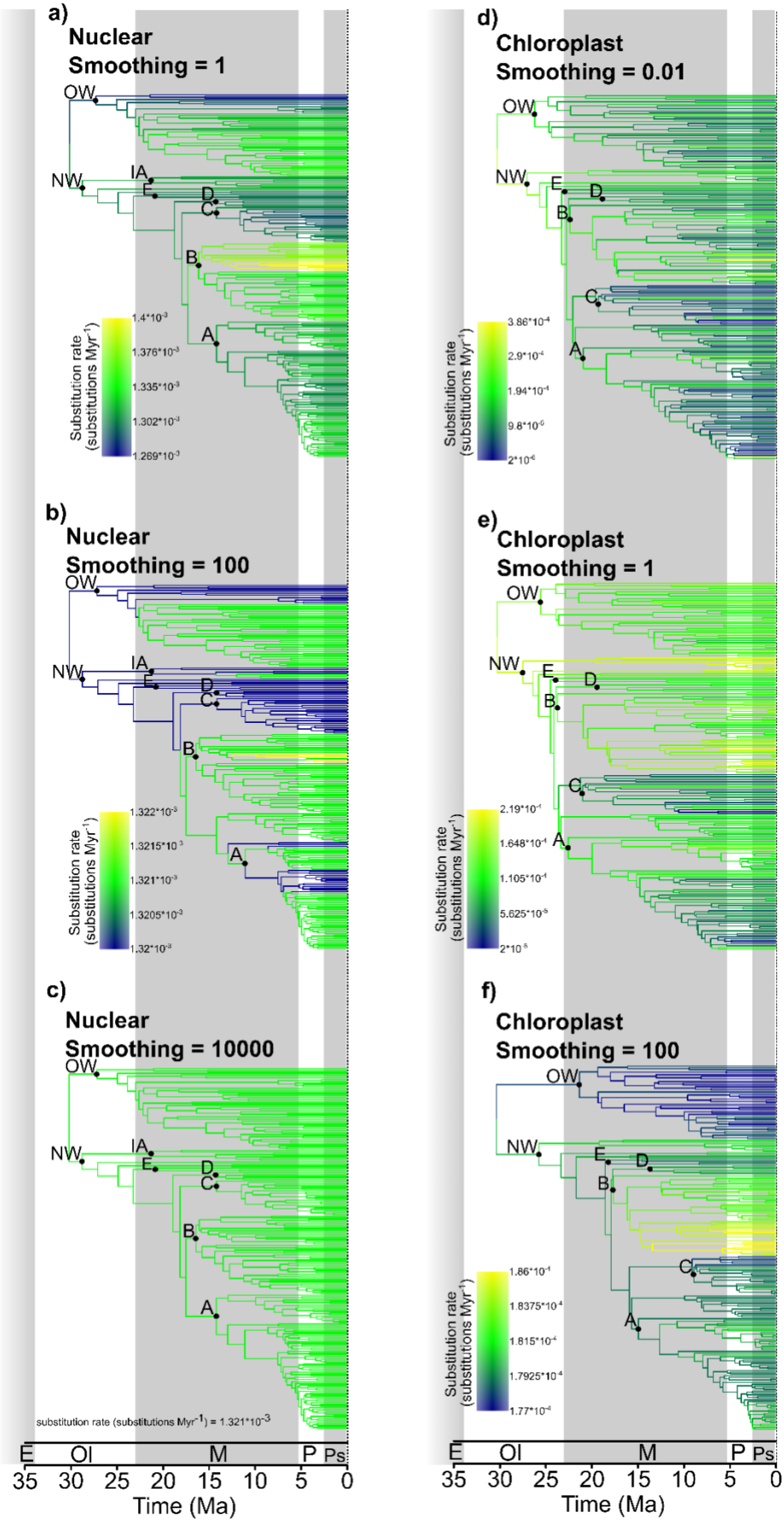

Figure A7.1: Time-calibrated phylogenies for *Ipomoea* inferred in treePL with alternative smoothing values from the optimum. Where rates differ between branches, branch colours refer to the inferred substitution rate. a) – c) show nuclear time-calibrated phylogenies inferred with smoothing values of 1, 100, and 10000 respectively. d) – f) show chloroplast time-calibrated phylogenies inferred with smoothing values of 0.01, 1, and 100 respectively.

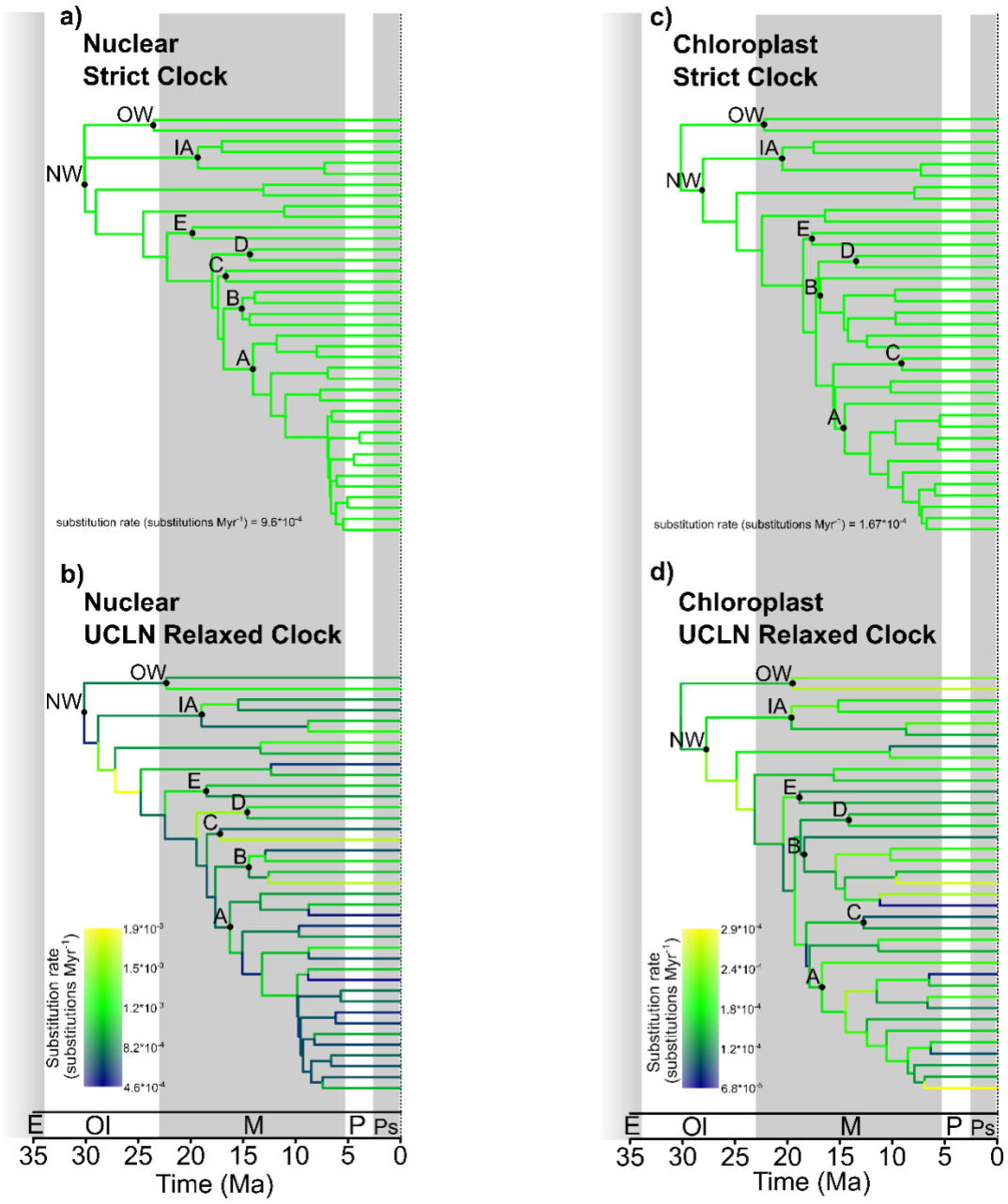

Figure A7.2. Skeletal time-calibrated phylogenies for *Ipomoea* inferred in RevBayes with either a strict clock, or a UCLN relaxed clock. Where rates differ between branches, branch colours refer to the inferred substitution rate. a) and b) are nuclear time-calibrated phylogenies inferred with either a strict clock a) or a UCLN relaxed clock b). c) and d) are chloroplast time-calibrated phylogenies inferred with either a strict clock c) or a UCLN relaxed clock d).

APPENDIX 8

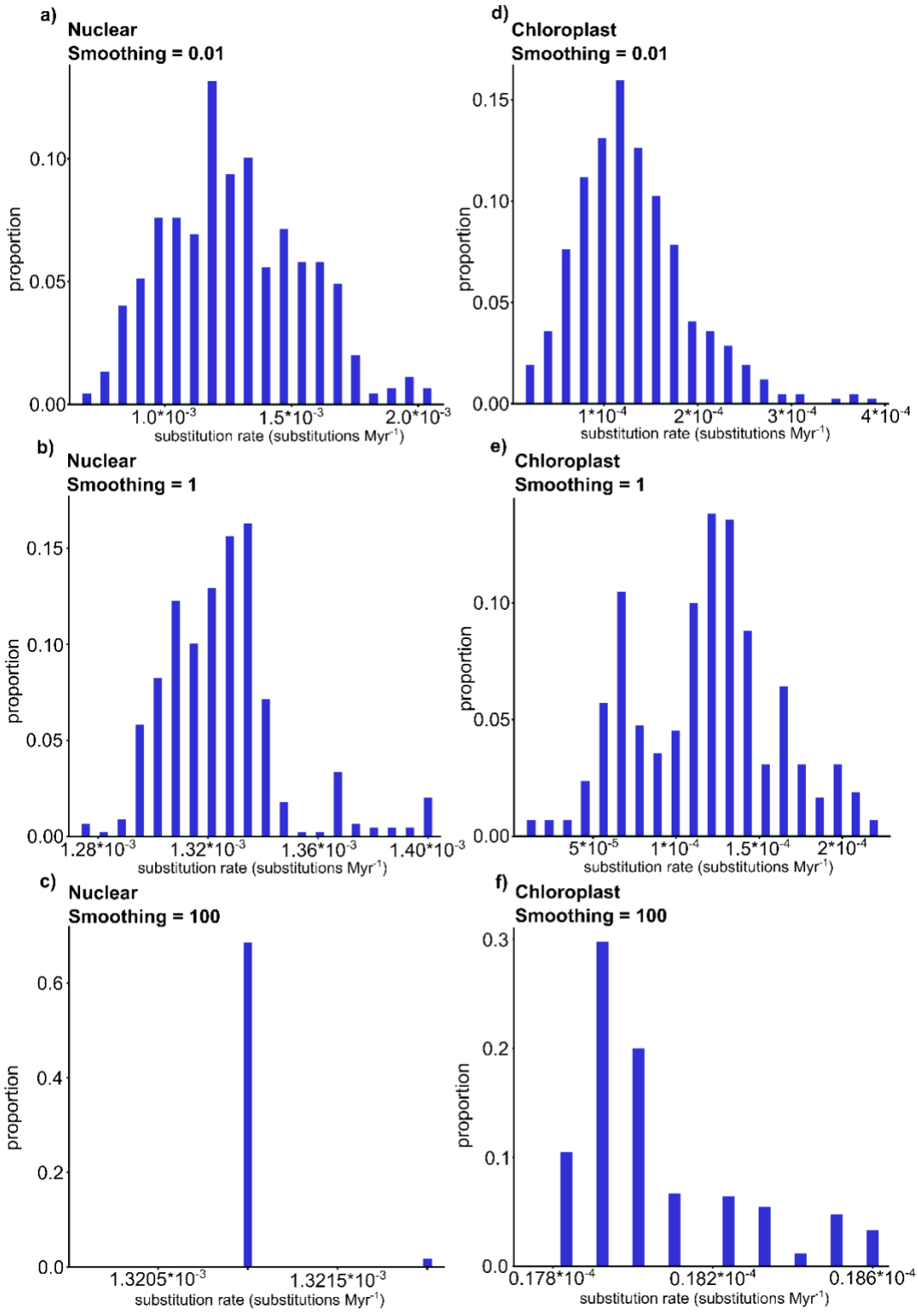

Figure A8.1. The distribution of inferred substitution rates in *Ipomoea* time-calibrated phylogenies when different smoothing values are used. The proportion and bin sizes are the same as set out in Figure A4.1. a) – c) show the distribution of substitution rates in nuclear time-calibrated phylogenies when using smoothing values of 0.01, 1, or 100. d) – f) show the distribution of substitution rates in chloroplast time-calibrated phylogenies when using smoothing values of 0.01, 1, 100. The inferred substitution rate when using a smoothing value of 10000 is not plotted because the same substitution rate is inferred for every branch.

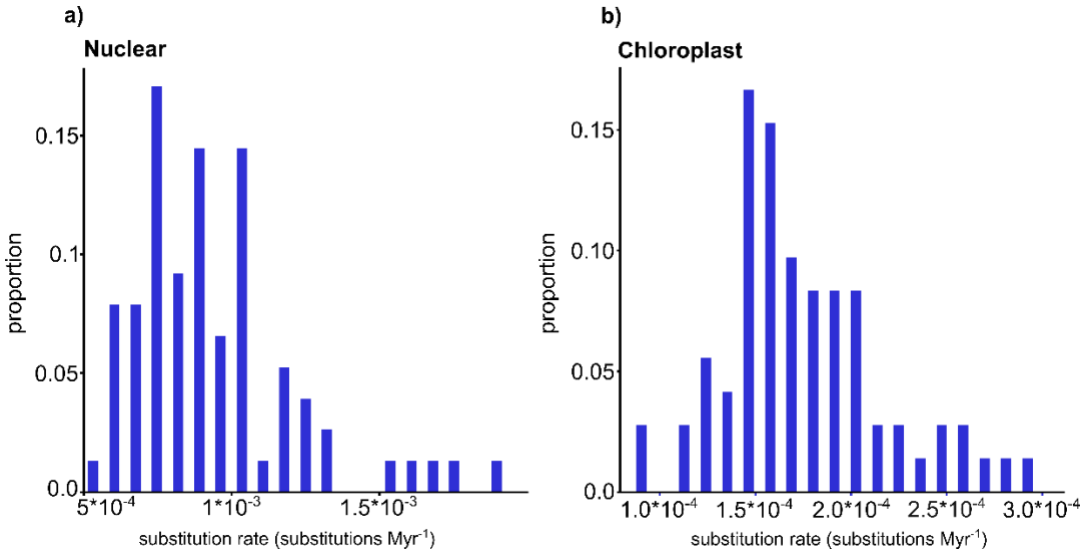

Figure A8.2. The distribution of inferred substitution rates in *Ipomoea* time-calibrated phylogenies inferred in RevBayes with a UCLN relaxed clock. The proportion and bin sizes are the same as set out in Figure A4.1. a) shows the distribution of substitution rates for the nuclear skeletal time-calibrated phylogeny. b) shows the distribution of substitution rates for the chloroplast skeletal time-calibrated phylogeny.

APPENDIX 9

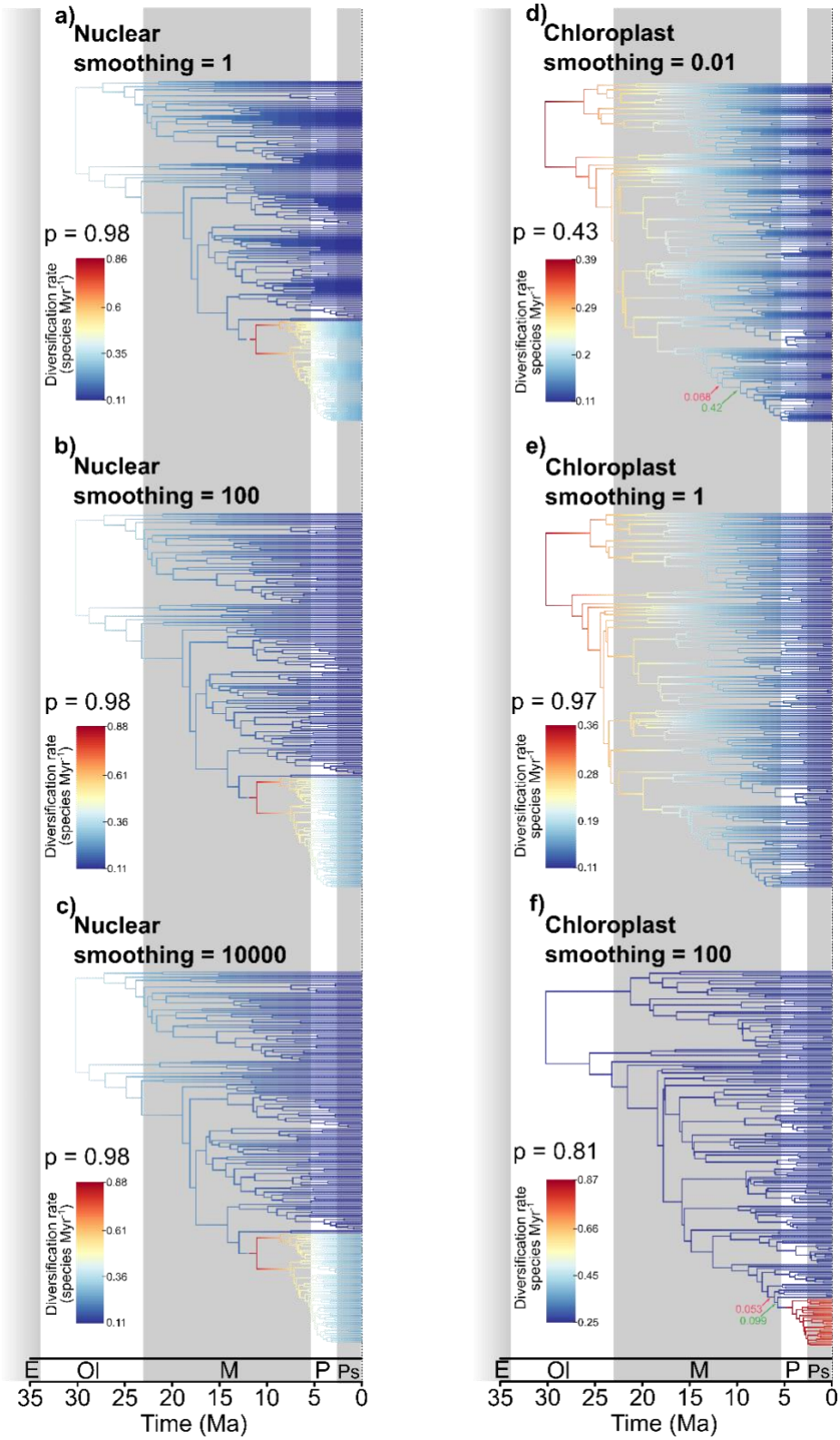

Figure A9.1. Net diversification rate estimates for *Ipomoea* inferred in BAMM with time-calibrated phylogenies inferred in treePL with smoothing values that differ from the optimum. Patterns of rate variation are indicated as set out in Figure 4.3. In a) – c) diversification parameters are estimated with nuclear time-calibrated phylogenies inferred with smoothing values of 1, 100, or 10000. In d) – f) diversification parameters are estimated with chloroplast time-calibrated phylogenies inferred with smoothing values of 0.01, 1, or 100.

APPENDIX 10

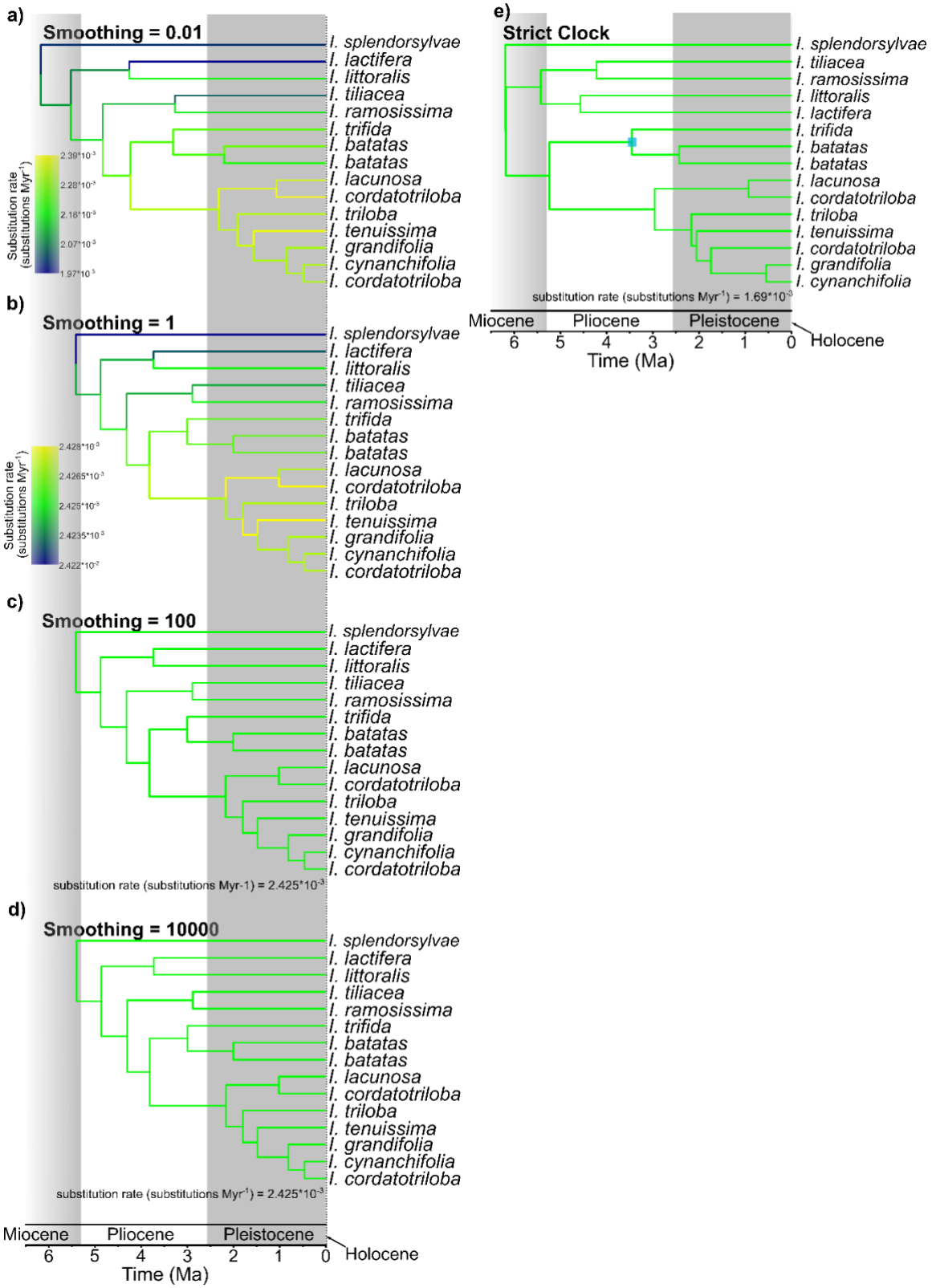

Figure A10.1. Nuclear time-calibrated phylogenies for *Ipomoea* sect. Batatas inferred with different assumptions about the magnitude of among-branch-substitution-rate variation. a) – d) show time-calibrated phylogenies for this clade inferred in treePL with different smoothing values. e) shows a time-calibrated phylogeny for this clade inferred in RevBayes with a strict clock. The blue bar refers to the 95% HPD for the divergence time between *I. batatas* and *I. trifida*.

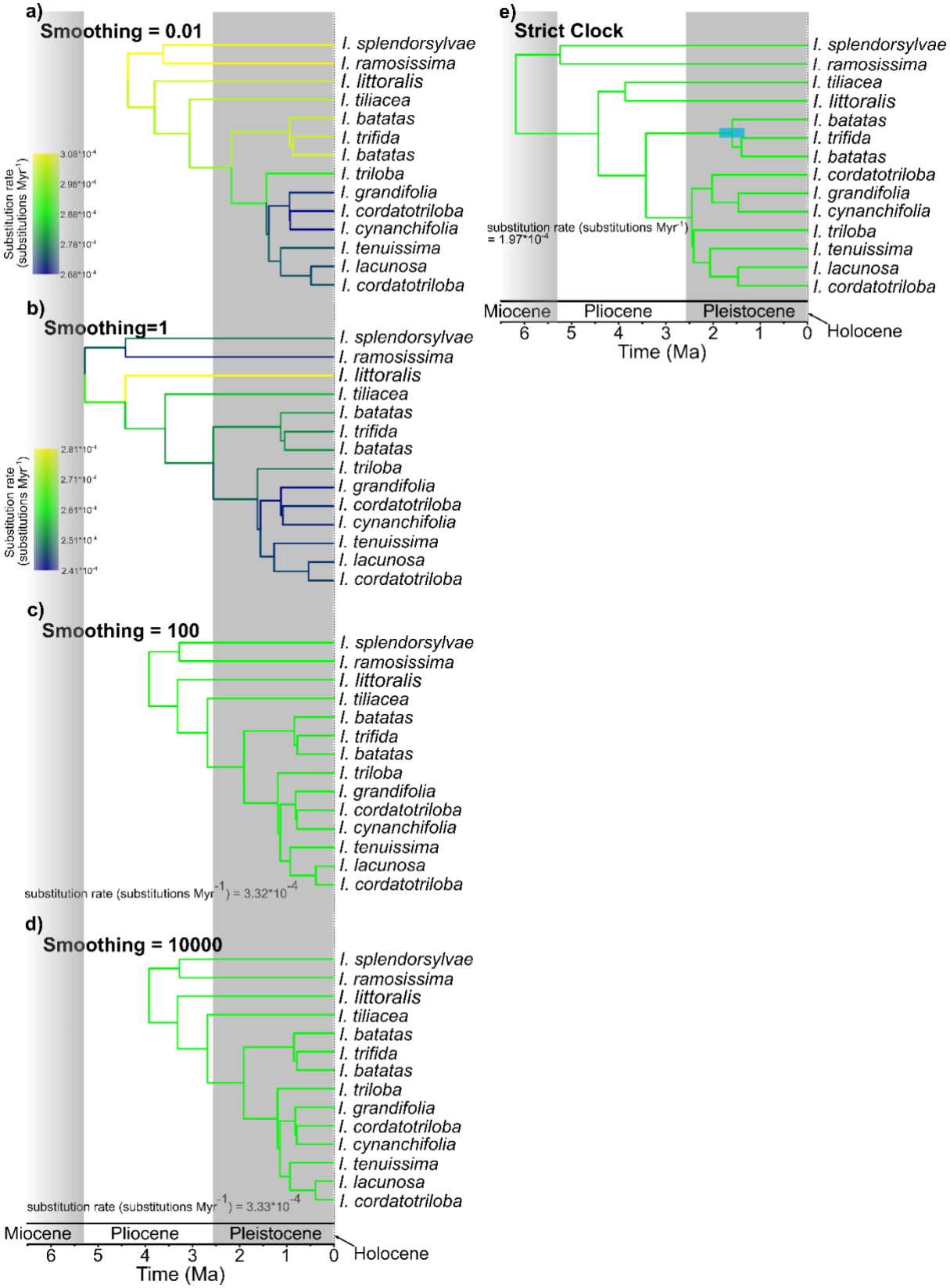

Figure A10.2. Chloroplast time-calibrated phylogenies for *Ipomoea* sect. Batatas inferred with different assumptions about the magnitude of among-branch-substitution-rate variation. a) – d) show time-calibrated phylogenies for this clade inferred in treePL with different smoothing values. e) shows a time-calibrated phylogeny for this clade inferred in RevBayes with a strict clock. The blue bar refers to the 95% HPD for the divergence time between *I. batatas* and *I. trifida*.

APPENDIX 11

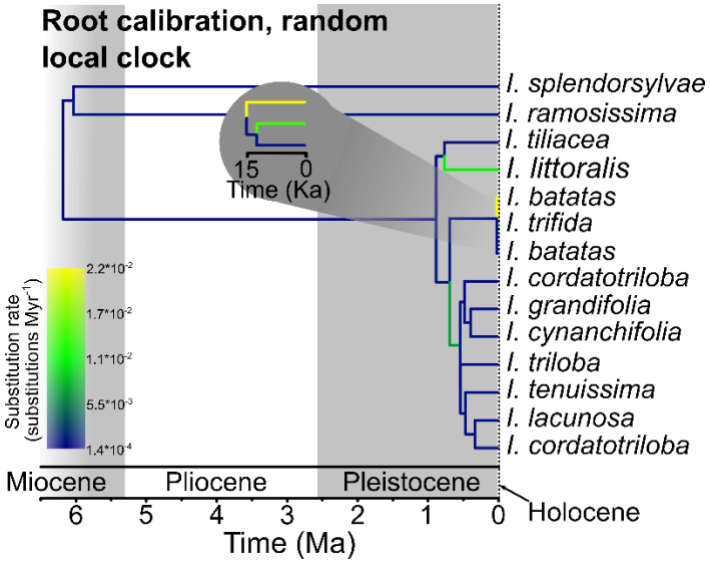

Figure A.11.1. An alternative chloroplast time-calibrated phylogeny for *Ipomoea* sect Batatas inferred in RevBayes. Branches are coloured according to the inferred substitution rate. Temporal calibrations are the same as in Figure 4.5a. A random local clock is used to infer substitution rates.

APPENDIX 12

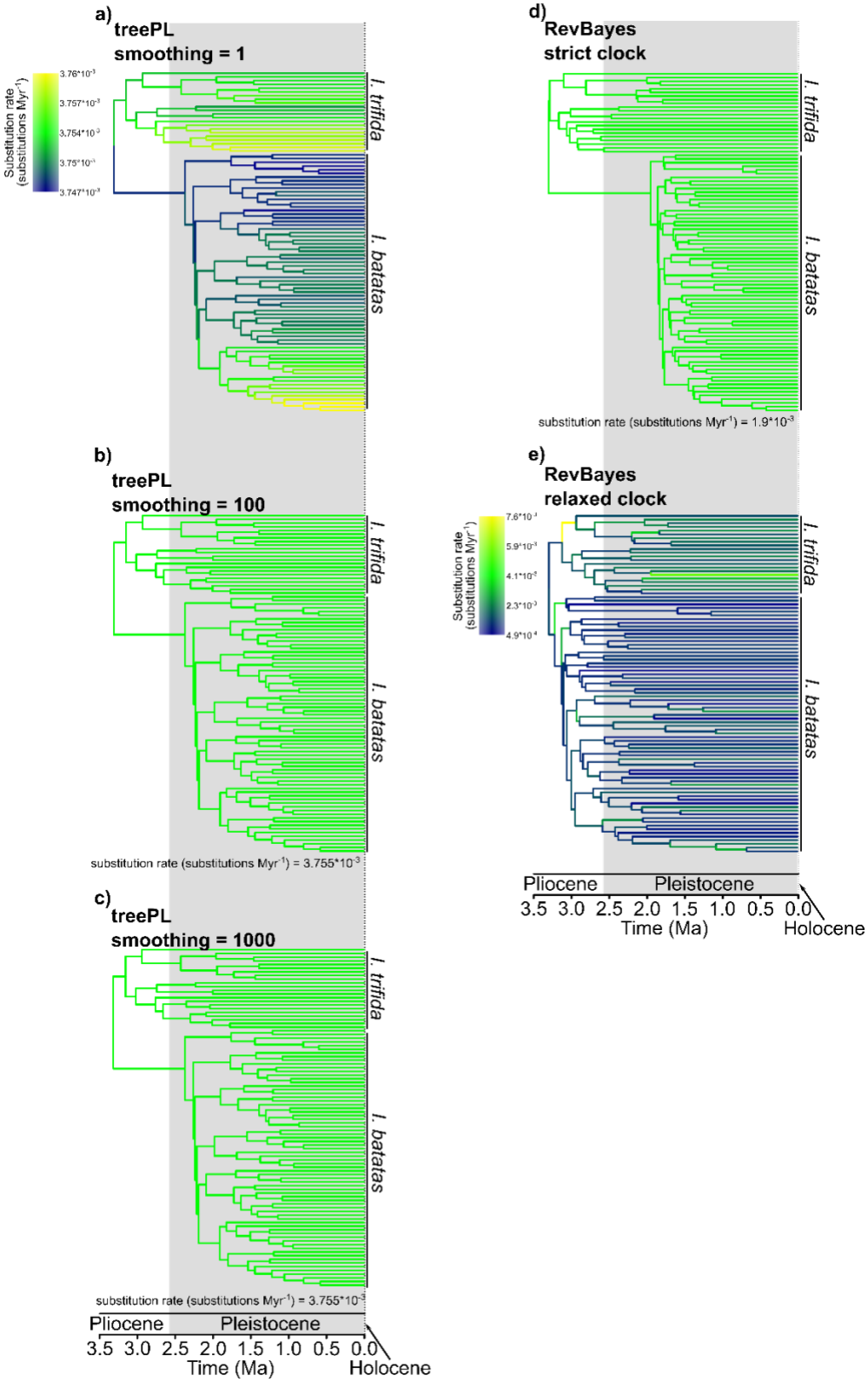

Figure A12.1. Nuclear time-calibrated phylogenies for all sampled accessions of *I. batatas* and *I. trifida* inferred with different assumptions about the magnitude of among-branch-substitution-rate variation. a) – c) show time-calibrated phylogenies inferred in treePL with different smoothing values. d) shows a time-calibrated phylogeny inferred in RevBayes with a strict clock, e) shows a time-calibrated phylogeny inferred in RevBayes with a UCLN relaxed clock.

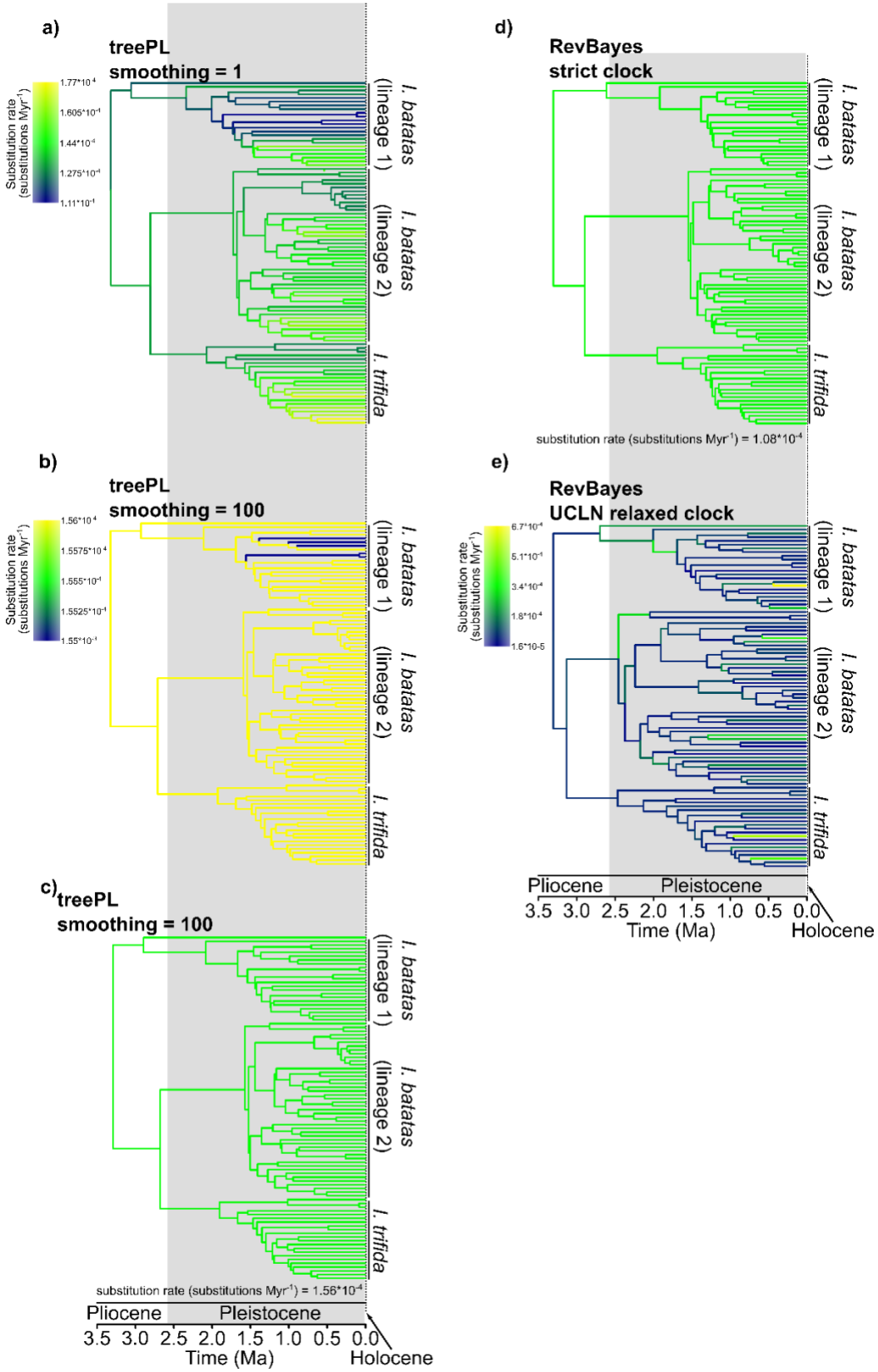

Figure A12.2. Chloroplast time-calibrated phylogenies for all sampled accessions of *I. batatas* and *I. trifida* inferred with different assumptions about the magnitude of among-branch-substitution-rate variation. a) – c) show time-calibrated phylogenies inferred in treePL with different smoothing values. d) shows a time-calibrated phylogeny inferred in RevBayes with a strict clock, e) shows a time-calibrated phylogeny inferred in RevBayes with a UCLN relaxed clock.

APPENDIX 13

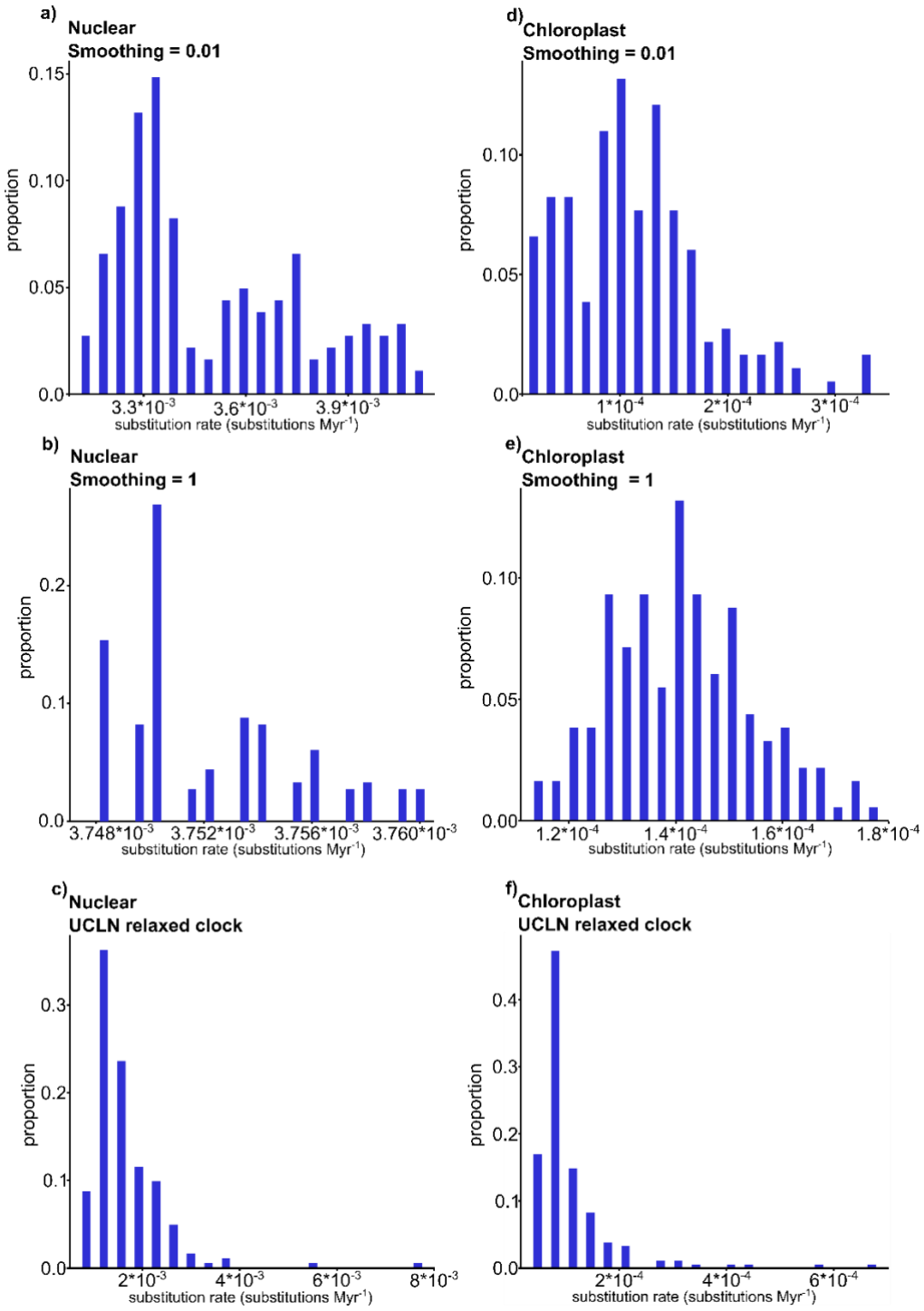

Figure A13.1. The distribution of inferred substitution rates in time-calibrated phylogenies for all sampled accessions of *I. batatas* and *I. trifida* with different assumptions about the magnitude of among-branch-substitution-rate-variation. The proportion and bin sizes are the same as set out in Figure A4.1. a) – c) show the distribution of substitution rates in nuclear time-calibrated phylogenies when using smoothing values of 0.01, 1, 100. d) shows the distribution of substitution rates in a nuclear time-calibrated phylogeny inferred with a UCLN relaxed clock. e) – h) show the distribution of substitution rates in chloroplast time-calibrated phylogenies when using smoothing values of 0.01, 1, 100. f) shows the distribution of substitution rates in a chloroplast time calibrated phylogeny inferred with a UCLN relaxed clock.

APPENDIX 14

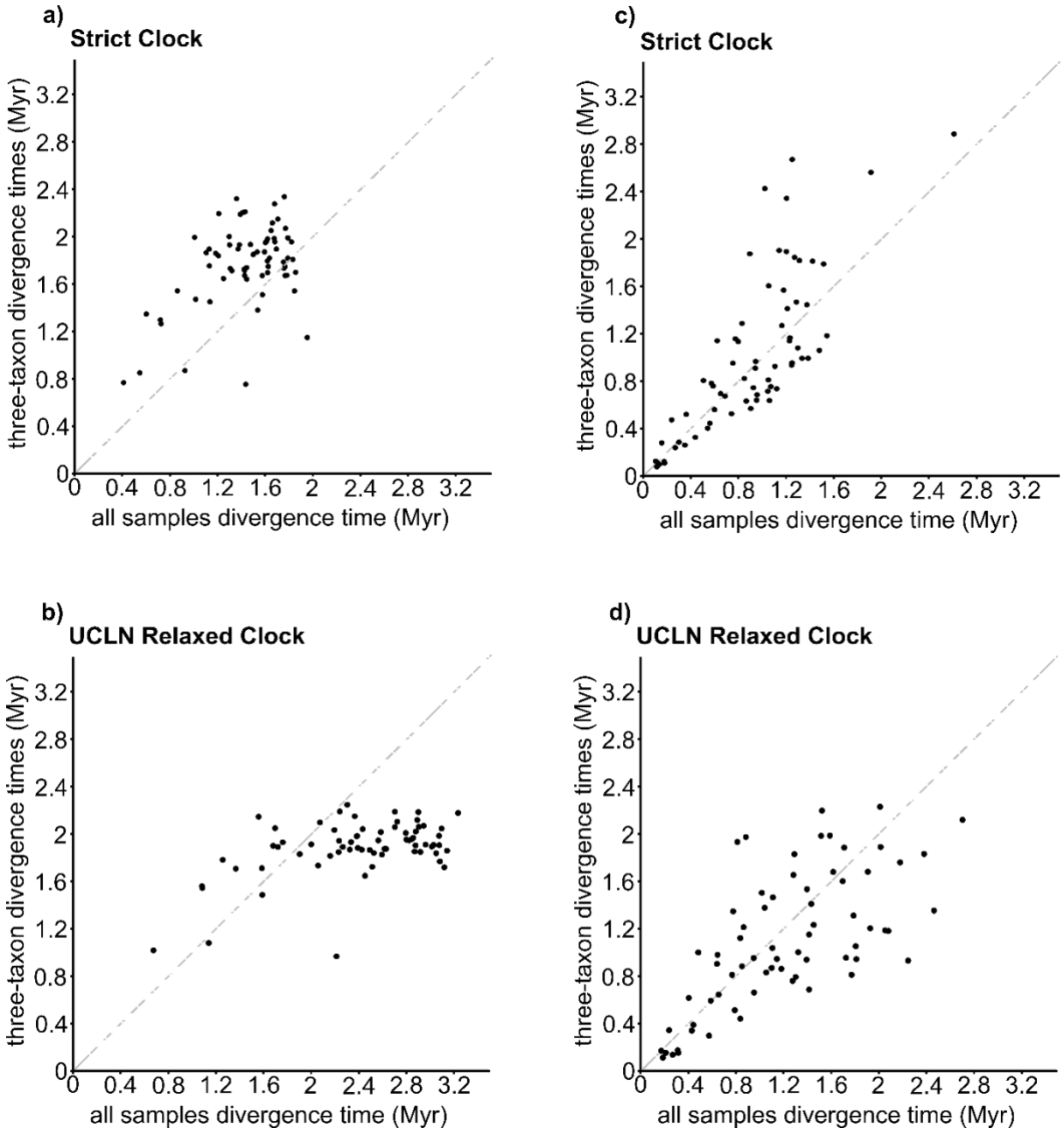

Figure A14.1. Comparison of divergence times between sampled lineages within *I. batatas* and *I. trifida*. Divergence times are estimated in either a time-calibrated phylogeny in which all sampled accessions are included, or a series of three-taxon time-calibrated phylogenies that sample equivalent nodes to those sampled in the time-calibrated phylogeny with all accessions. There are therefore two age estimates for each node, with the black plotted point in each graph referring to the MPE. In a) and b) time-calibrated phylogenies are inferred with nuclear data, and a strict clock a) or UCLN relaxed clock b). In c) and d) time-calibrated phylogenies are inferred with chloroplast data, and a strict clock c) or UCLN relaxed clock d).