Identifying apoplastic plant proteins involved in plant-pathogen interactions on the basis of their evolutionary profile



Felix Homma

Brasenose College

University of Oxford

Supervisor: Renier A.L. van der Hoorn

Thesis submitted for the degree of *Doctor of Philosophy*

September 2022

Abstract

To sustainably feed a growing global population, yield can be increased by reducing crop losses to plant pathogens. During infection of plants, several pathogens attempt to colonise the intercellular space, the apoplast. Apoplastic defence against attempted invasions includes the secretion of defence-related proteins, turning the apoplast into a molecular battlefield. While few apoplastic hydrolases have been characterised in plant immunity, many secreted proteins remain uncharacterised. Thus, in order to improve our understanding of apoplastic plant defence responses, we aimed to identify novel apoplastic proteins involved in immunity in this thesis. We combined comparative proteomics with molecular evolution and bioluminescence-based high-throughput disease assays. First, to identify core extracellular tomato proteins in immune responses, we studied apoplastic fluids (AF) of different biotic stress-induced tomato species *S.lycopersicum, S.peruvianum* and *S.pennellii* by mass spectrometry (MS). We identified that P69 subtilases were the most abundant proteins in tomato AFs during defence. Molecular evolution studies further implicated P69 subtilases in plant immunity and implicated another 20 biotic stress-induced tomato proteins under positive selection in plant immunity. Second, we compared apoplastic proteomic landscapes between tomato species and *N.benthamiana, G.max* and *A.thaliana* which marked distinct differences in profiles of accumulating defence proteins, implicating subtilases from *G.max* and chitinases from tomato species in plant immunity. Third, we established a bioluminescence-based, high-throughput disease assay for the *N.benthamiana-Pseudomonas syringae* pv. *tomato* (*Pto*) pathosystem, but failed to identify increased immunity in *N.benthamiana* against *Pto* for three proteins implicated in plant immunity, including P69B. Last, we used AlphaFold2 multimer as a high-throughput *in silico* screening tool to identify inhibitor-enzyme complexes from the plant-pathogen interface and identified 13 novel, putative hydrolase inhibitors, including six putative P69B inhibitors. This thesis has contributed to understanding apoplastic defence responses by identifying proteins implicated in plant immunity.

Acknowledgements

This thesis would have not been possible without the support and encouragement of colleagues, friends and family.

Firstly, I want to thank Prof. Renier van der Hoorn for the opportunity to be a member of the Plant Chemetics Laboratory for the past four years. I am grateful for your support, guidance and the academic freedom I enjoyed to explore biological questions and methodologies that captured my fascination. I appreciate the time and effort you spent to help me grow as a scientist and I enjoyed all of our scientific discussions.

Furthermore, I want to thank all Plant Chemetics Laboratory members I had the privilege to work with. Special thanks to Judith Paulus, Jiorgos Kourelis, Philippe Varennes-Jutras, Mariana Schuster, Marcel Bach-Pages and Riccardo Soldan; all of whom are inspiring scientists and great friends.

I want to thank my partner Anna Proffitt for her never-ending support, kindness and patience; your encouragement helped me to stay afloat during the most stressful of times.

I want to thank my family Inge, Norbert, Laura and Fabian Homma for their support and I am saddened that my late grandparents Hermann and Rosa Volz could not witness the end of my thesis. I want to thank the DTP and my fellow students Nicolas Arning, Niels Leadholm, Nasir Ahmad, Niels Wicke, Andreas Kjaer and Carolin Schulte, the Oxford University Basketball Club and Alex Bowmer, Daniel Thomson and Oskar Hoff for shaping these past four years.

Finally I want to thank the BBSRC, DTP, Department of Plant Sciences, Brasenose College and the Clarendon Fund for enabling my research.

Contents

1 Introduction **6**
- 1.1 Overview . . . 6
- 1.2 The apoplastic battlefield in plant immunity . . . 6
- 1.3 Pathogenesis-related proteins used in this thesis and the immuno secretome 8
- 1.4 Co-evolution and positive selection in plant immunity . . . 15
- 1.5 Structural prediction from co-evolution data . . . 18
- 1.6 Experimental implication in plant immunity . . . 19
- 1.7 Aims of the thesis: . . . 21

2 Tomato Immuno Secretomes **23**
- 2.1 Introduction . . . 23
- 2.2 Materials and Methods . . . 24
  - 2.2.1 Apoplastic Fluid extraction . . . 24
  - 2.2.2 Tomato genomic resources used . . . 25
  - 2.2.3 Estimating genetic distances between tomato species . . . 26
  - 2.2.4 Enzymatic characterisation of apoplastic fluid extracts . . . 27
  - 2.2.5 Activity-based protein profiling (ABPP) of serine hydrolases . . . 28
  - 2.2.6 Generating a *Solanum peruvianum* reference proteome . . . 28
  - 2.2.7 Preparing protein samples for mass spectrometry analysis . . . 29
  - 2.2.8 Mass spectrometry data processing and statistical analyses . . . 30
  - 2.2.9 Analysis of proteomics data of tomato apoplastic fluids . . . 30
  - 2.2.10 Ploidy analysis . . . 32
  - 2.2.11 Positive selection analysis . . . 32
- 2.3 Results . . . 34
  - 2.3.1 Apoplastic fluids from different tomato species . . . 34
  - 2.3.2 Generating a *Solanum peruvianum* reference proteome . . . 37
  - 2.3.3 Immuno secretomes of different tomato species . . . 39
  - 2.3.4 Positive selection in biotic stress-induced tomato secretomes . . . 49
- 2.4 Discussion . . . 59
  - 2.4.1 Protein concentration in wild tomato apoplastic fluid extracts increases upon Bion treatment . . . 60
  - 2.4.2 Assembling a genome and proteome of *Solanum peruvianum* . . . 61
  - 2.4.3 Comparing secretomes between different tomato species . . . 62
  - 2.4.4 Detecting molecular evolution in biotic stress-induced tomato secretomes . . . 64

3 Immuno Secretomes are different between plant taxons **68**
- 3.1 Introduction . . . 68
- 3.2 Materials and Methods . . . 70
  - 3.2.1 Plant growth conditions . . . 70
  - 3.2.2 Apoplastic Fluid extraction . . . 70
  - 3.2.3 Generating an unrooted species phylogenetic tree from proteomes . 71
  - 3.2.4 Generating a coding sequences-based phylogenetic tree . . . 71
  - 3.2.5 Enzymatic characterisation of apoplastic fluid extracts . . . 72
  - 3.2.6 Preparing protein samples for mass spectrometry analysis . . . 72
  - 3.2.7 Mass spectrometry data processing and statistical analyses . . . 73

3.2.8 Analysis of protein MS data of apoplastic fluid extracts 73
3.3 Results 75
3.3.1 Extracting biotic stress-induced apoplastic fluid extracts of different eudicot species 75
3.3.2 Preparing MS data for comparative proteomics 78
3.3.3 Comparative proteomics of six plant immuno secretomes 81
3.4 Discussion 98
3.4.1 Preparing MS data for comparative proteomics 98
3.4.2 Comparing PR protein accumulation between protein gels and MS data 98
3.4.3 Pareto orthogroups and their functional annotation 99
3.4.4 An emerging theme for apoplastic immunity proteins 101
3.4.5 Apoplastic defence proteomes induced by different elicitors 102
3.4.6 Outlook 103

4 AgromonasLux **105**
4.1 Introduction 105
4.2 Materials and Methods 107
4.2.1 Molecular cloning 107
4.2.2 AgromonasLux 109
4.2.3 Apoplastic fluid extraction 110
4.2.4 Western Blot analysis 111
4.2.5 Plate reader luminescence read out 111
4.2.6 Statistics 111
4.3 Results 112
4.3.1 Experimental design 112
4.3.2 Bioluminescent *Pto* 113
4.3.3 Effects of immune proteins on bioluminescence of *Pto in planta* . . 115
4.3.4 Test tomato candidate genes in AgromonasLux 120
4.4 Discussion 123
4.4.1 Experimental design 124
4.4.2 Establish AgromonasLux 125
4.4.3 Test tomato candidate proteins in agromonasLux 127

5 AlphaFold2 predicts protein complexes at the plant-pathogen interface **129**
5.1 Introduction 129
5.2 Materials and Methods: 133
5.2.1 Protein complex prediction with AlphaFold2 multimer: 133
5.2.2 Analysing output parameters of AlphaFold2 Multimer 134
5.2.3 Selection of tomato and pathogen proteomes and transcriptomes: . 134
5.2.4 Prediction of small, apoplastic proteins: 136
5.2.5 Visual scoring of candidate complexes: 137
5.2.6 RNAseq data mining, filtering and mapping: 137
5.2.7 Clustering of proteins of 37 plant pathogens in Orthogroups 139
5.3 Results 140
5.3.1 AlphaFold2 models experimentally validated protein complexes from the apoplastic plant-pathogen interface with high predicted quality 140
5.3.2 Modelling over 11,000 protein pairs with AlphaFold2 multimer . . . 148

5.3.3 Manual scoring of 376 protein complexes identifies putative novel apoplastic hydrolase inhibitors . . . 151
5.3.4 Expression of putative hydrolase inhibitors . . . 154
5.3.5 Conservation of putative hydrolase inhibitors . . . 155
5.3.6 Novel hydrolase inhibitors predicted by AlphaFold2 multimer . . . 157
5.4 Discussion . . . 167
5.4.1 AlphaFold2 models experimentally validated protein complexes at the apoplastic plant-pathogen interface with high predicted quality 167
5.4.2 Modelling over 11,000 protein pairs with AlphaFold2 multimer . . . 169
5.4.3 Manual scoring of 376 protein complexes identifies novel putative apoplastic hydrolase inhibitors . . . 171
5.4.4 Expression and conservation of putative hydrolase inhibitors . . . 172
5.4.5 Novel hydrolase inhibitors predicted by AlphaFold2 multimer . . . 173

6 General discussion and conclusions 177
6.1 Comparative mass spectrometry . . . 177
6.1.1 Generate apoplastic fluid extracts . . . 177
6.1.2 Comparing Bion-induced apoplastic proteomes . . . 178
6.2 Molecular evolution in the era of whole genome sequencing . . . 180
6.2.1 Genomics revolution and molecular evolution . . . 180
6.2.2 Pangenomics and molecular evolution . . . 181
6.3 Bioluminescent-based agromonasLux to test proteins in plant immunity . . . 181
6.4 Protein structure prediction to identify novel hydrolase inhibitors . . . 182
6.5 Conclusion . . . 184

1 Introduction

1.1 Overview

A growing global population places agriculture under tremendous stress to sustainably produce enough food without expanding agricultural land (van Dijk et al. 2021). One approach to increase yield is to reduce losses to pathogens and disease which amounts to 8.1-41.1% annualy for major crops (Savary et al. 2019). To control pests, pesticides and resistance breeding are used. As continuous application of the same pesticides resulted in pesticide-resistant pathogens, breeding disease resistant plants to increases their natural ability to fend off pathogens is a complementary approach (Lucas et al. 2015). Single-locus resistance breeding provides strong, but often temporary immunity as evolutionary pressures on pathogens to evade a single resistance gene are high (Iida et al. 2015). Alternatively, durable genetic disease resistance can be achieved by stacking single-locus resistance genes or breeding for broad-spectrum, quantitative disease resistance (Nelson et al. 2018; Crété et al. 2020). To understand durable broad-spectrum disease resistance, we need to identify and characterise proteins involved in plant immunity.

1.2 The apoplastic battlefield in plant immunity

In ongoing competition for limited resources, microbial plant pathogens attempt to colonise plants, extract nutrients, multiply and propagate. To fend off pathogens, plants have pre-formed defence mechanisms such as physical barriers (cuticules and cell walls) and antimicrobial chemicals (phytoanticipins) as well as inducible defence responses (Serrano et al. 2014; Wan et al. 2021). Inducible defence responses differ depending on the infecting pathogen's lifestyle, which can be divided into necrotrophic and biotrophic (Glazebrook 2005). Biotrophs feed off living tissue while necrotrophs feed of dead tissue and hemibiotrophs combine both strategies throughout their life cycle. Defence against pathogens with such distinct life styles requires distinct plant defence responses, induced and regulated by different plant stress hormones. Against necrotrophic pathogens,

plants generate jasmonic acid- (JA) and ethylene- (ET) induced systemic resistance (ISR) while mounting salicylic acid- (SA) dependent systemic acquired resistance (SAR) against biotrophs (Thomma et al. 1998). Biotrophs and hemibiotrophs often attempt to colonise the plant intercellular space, the apoplast. The plant defends the apoplast with a multi-layered response to prevent hostile takeover and pathogens attempt to suppress and evade plant defence responses. This turns the intercellular space into a molecular battlefield (Doehlemann and Hemetsberger 2013).

The most widely cited model describing plant immunity against biotrophs is the 'zig-zag-zig' model, dividing antagonistic plant-pathogen interactions into three different levels: First, the plant recognises general damage-, microbe- and pathogen- associated molecular patterns (D/M/PAMPs) triggering pattern-triggered immunity (PTI). PTI is triggered through pattern-recognition receptors (PRRs) residing in the plasma membrane. PRRs are most often leucine-rich repeat receptor-like proteins (RLPs) or receptor-like kinases (RLKs). Second, pathogens deploy secreted effectors suppressing PTI, resulting in effector-triggered susceptibility (ETS). Third, plants evolved to recognise effectors, resulting in effector-triggered immunity (ETI) which culminates in the hypersensitive response (HR). ETI is described to be mediated by intracellular resistance proteins which are most often nucleotide-binding, leucine-rich repeat receptors (NLRs). The zig-zag-zig model provides a conceptual framework based on the recognised molecules triggering the immune response (Jones and Dangl 2006; Ngou et al. 2022). Recently, evidence has blurred the lines between PTI and ETI as both share signalling components and potentiate each other (Ngou et al. 2021; Yuan et al. 2021). An alternative 'spatial immunity model' divides plant immunity into intracellularly- and extracellularly triggered immunity (InTI and ExTI), mediated by NLRs and RLPs/RLKs respectively (van der Burgh and Joosten 2019). Regardless of preferred models to frame plant immunity, RLPs, RLKs and NLRs are central receptors in perceiving pathogenesis-related molecules to trigger defence responses in primary infected tissue (Wiermer et al. 2005; Kong et al. 2016; Zhang and Li 2019; Ngou et al. 2022). These responses include the increased accumulation of reactive oxygen species (ROS), activation of mitogen-activated protein kinases (MAPKs), tran-

scriptional reprogramming, production of phytoalexins and accumulation of salicylic acid (SA), increasing local disease resistance and inducing systemic acquired resistance (SAR) (Pieterse et al. 2009; Vlot et al. 2009; Seyfferth and Tsuda 2014; Stael et al. 2015; Klessig et al. 2018). SAR is the phenomenon of prolonged, increased and broad-spectrum systemic disease resistance against biotrophic pathogens of previously uninfected leaves. This was first described in tobacco (*Nicotiana tabacum*) plants with increased disease resistance to secondary *Tobacco mosaic virus* (TMV) infections in previously uninfected leaves (Ross 1961). SAR was linked to SA (White 1979) because transgenic plants depleted of SA show reduced SAR (Gaffney et al. 1993; Delaney et al. 1994; Lawton 1995). Resistance against biotrophic pathogens could be increased by exogenous application of SA or its functional analogue benzothiadiazole S-methyl ester (BTH) (**Figure 1**) (Lyr et al. 1999; Navarre and Mayo 2004; Vlot et al. 2009; Takatsuji 2014). SA directly interacts with a plethora of different proteins including the master regulator of SA-mediated immunity, the *non-expressor of PR gene 1* (NPR1) (Cao et al. 1994; Delaney et al. 1995; Glazebrook et al. 1996; Shah et al. 1997). Npr1-mediated expression of defence-related genes is is suppressed by antagonistically acting SA receptors NPR3/4 (Wu et al. 2012; Manohar et al. 2015; Ding et al. 2018). High levels of SA, however, inhibit the suppression of defence-related gene expression through NPR3/4 and promote expression of defence-related genes via NPR1 (Zhang and Li 2019). In agriculture, SA and functional analogues are used to prime broad-spectrum disease resistance (Dempsey and Klessig 2017). One commercially available product containing BTH as the active ingredient is Syngenta's Bion.

1.3 Pathogenesis-related proteins used in this thesis and the immuno secretome

A hallmark of SAR is high accumulation of pathogenesis-related (PR) proteins in the apoplast and vacuole (Van Loon and Van Kammen 1970; Gianinazzi and Kassanis 1974). PR proteins are divided into 17 groups and are defined as proteins highly accumulating upon biotic stress trigger (van Loon et al. 2006). Initial studies of SAR focused on tobacco (*Nicotiana tabacum*) and *Arabidopsis thaliana* and found SAR against biotrophic

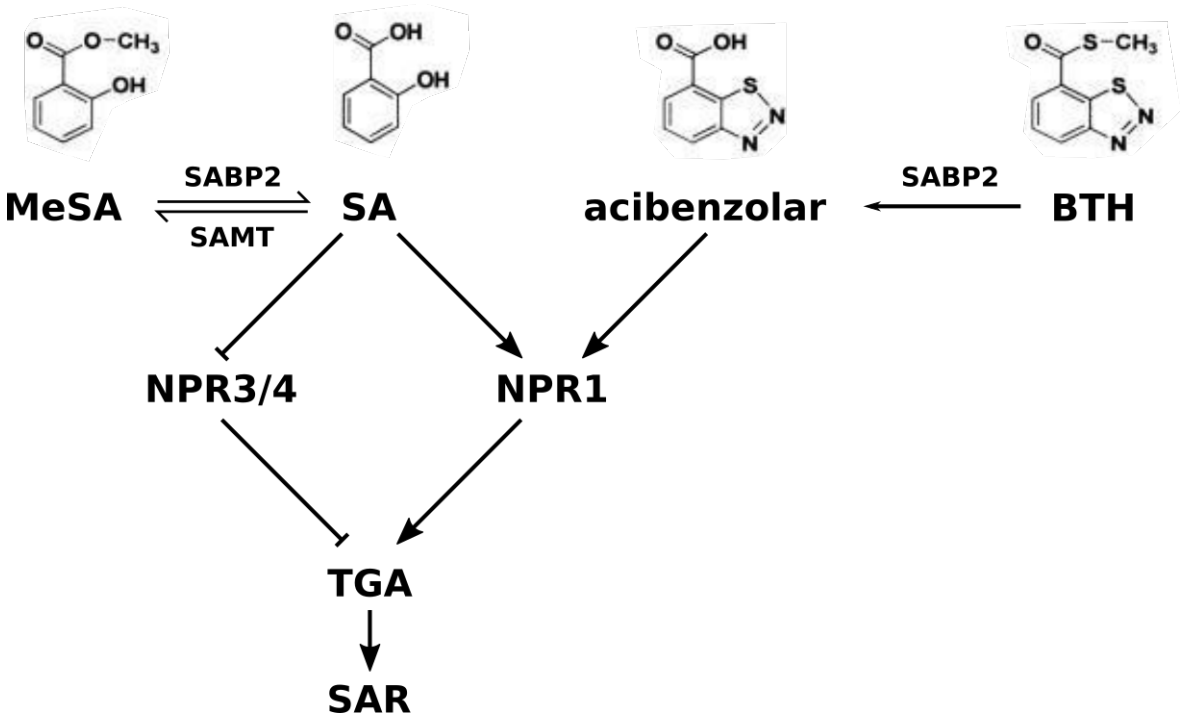


Figure 1: Convergence of SA and BTH in NPR1-mediated systemic acquired resistance (SAR). The recognition of a biotrophic pathogen results in local accumulation of salicylic acid (SA). High accumulation of SA results in conversion of SA to methylated SA (MeSA) by SA-methyl transferase (SAMT). MeSA as well as BTH can travel to distant tissue where they are converted to SA and acibenzolar by SA-binding protein 2 (SABP2) respectively. SA directly binds to NPR1, conferring TGA-mediated SAR and acibenzolar-mediated SAR is NPR1-dependent. SA furthermore binds to NPR3/4, suppressing the inhibition of TGA-mediated SAR. BTH is the active ingredient in the commercial pesticide Bion. The diagram was taken and modified from (Tripathi et al. 2010)

pathogens associated with apoplastic accumulation of PR1, PR2 (β-1,3-glucanase), PR3 (chitinases) and PR5 (thaumatin-like protein). (Van Loon and Van Kammen 1970; Gianinazzi and Kassanis 1974; Parent et al. 1985; Jamet and Fritig 1986; Legrand et al. 1987; Uknes et al. 1992). Further studies of SAR revealed increased apoplastic peroxidase activity (PR9) and high accumulation of PR7 (P69 subtilases) in tomato (Van Loon and Antoniw 1982; Tornero et al. 1996).

PR1 proteins belong to the cysteine-rich secretory proteins, antigen 5, and pathogenesis-related-1 (CAP) protein family (Breen et al. 2017). Plant genomes encode multiple CAPs (13 in tomato, 22 in *A.thaliana* and 32 in rice *Oryza sativa*) and these include vacuolar and secreted proteins (van Loon et al. 2006; Akbudak et al. 2020). Members of the PR1 protein family confer increased disease resistance against bacteria (Sarowar et al. 2005;

Shin et al. 2014), oomycetes (Alexander et al. 1993; Sarowar et al. 2005), fungi (Kiba et al. 2007), but not viruses (Cutt et al. 1989) when over expressed *in planta* and only one PR1 protein increases in abundance upon biotic stress in *A.thaliana* (At2g14610) (van Loon et al. 2006; Breen et al. 2017). Recently, PR1 proteins were shown to bind sterol at low physiological concentrations ($K_d = 28\mu M$) (Gamir et al. 2017). This supported the hypothesis that direct anti-oomycete effects of PR1 against sterol-auxotroph pathogens such as *Phytophthora infestans* could stem from PR1-mediated sequestering of sterol (Gamir et al. 2017). Furthermore, PR1 proteins encode a conserved, embedded CAP-derived peptide (CAPE) that can increase resistance against biotrophic *Pseudomonas syrinage* pv. *tomato* DC3000 in *A.thaliana* (Chen et al. 2014). The CAPE peptide is embedded in the sterol-binding region and PR1 is hypothesised to function as a sterol sensor, releasing an immunogenic DAMP peptide after sterol binding (Breen et al. 2017). PR1 proteins are targeted by effectors of necrotrophs, promoting their virulence and likely hijacking a defence mechanism against biotrophic pathogens (Breen et al. 2017).

PR2 (β-1,3-glucanases) are small (33-36 kDa) beta-glucanases in plants, that can hydrolyse β-1,3-glucans. They comprise five different structural groups (Doxey et al. 2007; Sinha et al. 2014; Xu et al. 2016) that all contain an N-terminal secretion signal and the core glycosyl hydrolase family 17 domain, while groups I and IV have an additional hydrophobic C-terminal extension, which are possibly vacuolar targeting sequences (Doxey et al. 2007). Plant genomes encode many β-1,3-glucanases (27 in rice and 50 *A.thaliana* and tomato each). (Doxey et al. 2007; Paniagua et al. 2021; Perrot et al. 2022) that have been implicated in physiological processes such as embryogenesis (Helleboid et al. 1998), seed germination (Leubner-Metzger et al. 1995), cold tolerance (Yaish et al. 2006; Romero et al. 2008) and plant defence (Balasubramanian et al. 2012). PR2s group II β-1,3-glucanases provide increased resistance against a broad range of fungi and oomycetes (van Loon et al. 2006; Doxey et al. 2007; Balasubramanian et al. 2012), likely by regulating callose turnover and release of immunogenic elicitors. In *A.thaliana*, two PR2 proteins (At3g57240 and At3g57260) accumulate after SA treatment (Datta and Muthukrishnan 1999; Doxey et al. 2007). Plant pathogens in turn evolved to secrete glucanase inhibiting

proteins (GIPs) to suppress hydrolytic and antimicrobial activity of PR2s (Ham et al. 1997; Damasceno et al. 2008). *Phytophthora sojae* abundantly secretes GIPs, which cover its mycelial surface in a protective layer of protein armour, highlighting pathogen adaptation to the hydrolytic apoplastic battlefield (Rose et al. 2002). Over expressing PR2 in crops such as rice, tomato and tobacco increases resistance against filamentous pathogens, which can be further increased by co-expression with other PR proteins, such as PR3 and PR5 (Balasubramanian et al. 2012; Boccardo et al. 2019).

PR3 are small (25-35 kDa) glycosyl hydrolases degrading chitin (Flach et al. 1992; Sinha et al. 2014). Chitinases were identified as highly-accumulating PR proteins in the tobacco apoplast during SAR (Parent et al. 1985; Jamet and Fritig 1986; Legrand et al. 1987) and they can be divided into class I, II and IV chitinases. Class I and IV chitinases contain an N-terminal, cysteine-rich chitin-binding domain which is absent from class II chitinases (Collinge et al. 1993; Araki and Torikata 1995). Chitinases degrade chitin, which is a β-1,4-linked polymer of N-acetylglucosamine and a major component of insect exoskeletons and fungal cell walls (Ravi Kumar 2000). Chitinases have *in vitro* and *in vivo* anti-fungal activity (Broglie et al. 1991; Collinge et al. 1993; Grison et al. 1996) and degradation of chitin impairs structural integrity of fungal hyphae and releases elicitors, triggering further plant immune responses (Schlumbaum et al. 1986; Takashima et al. 2020). Plant pathogens in turn have evolved an array of mechanisms to evade chitinase-mediated immunity, including protease-mediated degradation of plant chitinases and sequestering of released immunogenic elicitors (Buscaill and van der Hoorn 2021).

PR5 proteins are small (21-26 kDa), cysteine-rich proteins, first described to accumulate in apoplastic fluid of tobacco (*N.tabacum*) upon tobacco mosaic virus (TMV) infection (Cornelissen et al. 1986; Liu et al. 2010). PR5 belong to the protein group of thaumatin-like proteins (TLPs), based on sequence similarity to a sweet-tasting protein from *Thaumatococcus daniellii*; no PR5 proteins are reported to taste sweet, however (van der Wel and Loeve 1972; Liu et al. 2010). TLPs are either targeted for secretion into the extracellular space or into the vacuole. Plant genomes encode multiple TLPs (31 in rice and 50 in *A.thaliana*) (Anžlovar and Dermastia 2003; Liu et al. 2010) that consist of three

distinct domains which are stabilised by up to eight disulfide bridges. Domains I and II form a cleft structure which can be either acidic, neutral or basic (Min et al. 2004; Ghosh and Chakrabarti 2008). The charge of the cleft region is associated with different binding properties of TLPs (Liu et al. 2010). Binding targets include actin (Takemoto et al. 1997), ice crystals (Yu and Griffith 1999) and G-protein-coupled receptors (such as the human receptor for sweet) (Li et al. 2002; Liu et al. 2010) as well as β-1,3-glucans (Trudel et al. 1998), cell wall mannoproteins (Ibeas et al. 2000) and xylanases (Fierens et al. 2007). PR5 TLPs have anti-fungal and anti-oomycete activity and several modes of actions have been proposed. These include binding fungal cell wall components and interfering with hyphal growth or recruiting PR2 β-1,3-glucanases to hydrolyse β-glucans, as well as integration of PR5 into the fungal membrane, forming membrane-permealising pores (Roberts and Selitrennikoff 1990; Trudel et al. 1998; Grenier et al. 1999; Liu et al. 2010). Some TLPs exhibit β-1,3-glucanase activity themselved, which was not strictly associated with anti-fungal properties (Barre et al. 2000; Menu-Bouaouiche et al. 2003). PR5 has been over expressed in transgenic crops such as strawberry, wheat and tobacco to increase broad-spectrum disease resistance against fungi and oomycetes (Liu et al. 2010). Next, P69 subtilases (PR7) are large (69 kDa), secreted endopeptidases and an expansion within the SBT1.10 subtilase clade in *Solanaceae* species (Taylor and Qiu 2017). PR7s were first described in tomatoes and highly accumulate upon biotic stress trigger in tomato (*Solanum lycopersicum*), but not in wild tobacco (*Nicotiana benthamiana*) (Tornero et al. 1996; Tian et al. 2004; Grosse-Holz et al. 2018; Planas-Marquès et al. 2018). Based on a recent family tree of tomato, P69 subtilases encode eleven different members (A-K) (Reichardt et al. 2018). Confusingly, naming of P69 subtilases in the literature is not consistent and we will use names assigned to Sol genomic protein identifiers in Reichardt et al. 2018. Upon biotic stress, P69A, B, D, G and J highly accumulate in tomato apoplastic fluid (Planas-Marquès et al. 2018). Surprisingly, no direct antimicrobial activity of any PR7 has been detected yet, but P69B activates the immune papain-like cysteine protease (PLCP) Rcr3 by cleaving off its inhibitory pro domain (Paulus et al. 2020). Furthermore, P69B specifically releases an elicitor peptide from the *P.infestans* effector PC2 (Wang

et al. 2021a). *P.infestans* in turn, secretes two kazal-like inhibitors, Epi1 and Epi10, suppressing hydrolytic activity of P69B and highlighting pathogen adaptation to highly abundant plant proteins in the apoplast of defence-induced plants (Tian et al. 2004; Tian et al. 2005; Tian and Kamoun 2005).

PR9 proteins belong to plant-specific, secreted class III haem peroxidases which accumulate in the apoplast upon biotic stress (Van Loon and Antoniw 1982). Plant genomes encode different class III peroxidases (138 in rice and 73 in *A.thaliana*) (Tognolli et al. 2002; Passardi et al. 2004), involved in different physiological processes including lignification (Sasaki et al. 2004), root elongation (Passardi et al. 2006), germination (Bailly et al. 2004), salt tolerance (Llorente et al. 2002) and plant defence (Choi et al. 2007; Daudi et al. 2012; Shigeto and Tsutsumi 2016). Functionally, class III peroxidases either oxidise substrates such as lignin precursors during lignification, or regulate ROS levels (Chen and Schopfer 1999; Sasaki et al. 2004; Choi et al. 2007; Daudi et al. 2012). Silencing of two *A.thaliana* peroxidases increased susceptibility to bacterial and fungal pathogens while impairing the oxidative burst (Daudi et al. 2012) and over expression of a rice peroxidase in carrot *Daucus carota* conferred increased disease resistance against *Alternaria radicina* and reduced ROS burst (Wally and Punja 2010). PR9s thus act in plant defence in a ROS generating and scavenging role.

PR1, 2, 3, 5, 7 and 9 are markers of SAR in different species due to their high accumulation upon SA-dependent biotic stress signalling and are summarised in **Table 1**. In addition to SAR markers, other PR proteins can be induced during defence responses against biotrophic pathogens. These include more chitinases (PR4, PR8 and PR11), tomato protease inhibitor I (PR6), cytosolic ribonuclease-like proteins (PR10), defensins (PR12), thionins (PR13), lipid transfer proteins (PR14), oxalate oxidases (PR15), oxalate-like oxidases (PR16) and PRp27 (PR17) (van Loon et al. 2006). Of these, class III chitinases (PR8) and their dual chitinase and lysozyme enzymatic activity might act against fungi and bacteria while defensins (PR12), thionins (PR13) and some lipid transfer proteins (PR14) have antifungal, antioomycete and anti-bacterial effects (van Loon et al. 2006).

More proteins, not classified as PR proteins, increase in abundance in the extracellu-

Table 1: Summary of PR proteins marking systemic acquired resistance

ID	Functional annotation	Description
PR1	cysteine-rich secretory proteins, antigen 5, and pathogenesis-related-1 (CAP)	Sterol-binding protein (Gamir et al. 2017) with embedded immunogenic CAP-derived peptide (Chen et al. 2014; Breen et al. 2017). Increase disease resistance against bacteria (Sarowar et al. 2005; Shin et al. 2014), oomycetes (Alexander et al. 1993; Sarowar et al. 2005) and fungi (Kiba et al. 2007)
PR2	β-1,3-glucanases	Hydrolysis of β-1,3-glucans in fungal and oomycete cell walls, damaging structural integrity and releasing immunogenic elicitors (Balasubramanian et al. 2012). Increase disease resistance against fungi and oomycetes (van Loon et al. 2006; Doxey et al. 2007; Balasubramanian et al. 2012)
PR3	class I, II and IV chitinases	Glycosyl hydrolases degrading chitin, (Flach et al. 1992) damaging structural integrity and releasing immunogenic elicitors (Schlumbaum et al. 1986; Takashima et al. 2020). Increase disease resistance against fungi (Broglie et al. 1991; Collinge et al. 1993; Grison et al. 1996)
PR5	Thaumatin-like proteins	Diverse binding properties, targeting actin (Takemoto et al. 1997), β-1,3-glucans (Trudel et al. 1998), cell wall mannoproteins (Ibeas et al. 2000) and xylanases (Fierens et al. 2007). Increase disease resistance against fungi (Liu et al. 2010), oomycetes (Liu et al. 2010)
PR7	P69 subtilases	Endopeptidases with expansion within SBT1.10 subtilase clade in *Solanaceae* species (Taylor and Qiu 2017). P69B processes preRcr3 (Paulus et al. 2020) and releases immunogenic peptide from *Phytophthora infestans*-derived effector PC2 (Wang et al. 2021b).
PR9	class III haem peroxidases	Peroxidases involved in lignification (Sasaki et al. 2004) and plant defence (Choi et al. 2007; Daudi et al. 2012; Shigeto and Tsutsumi 2016). Increase disease resistance against fungi (Wally and Punja 2010; Daudi et al. 2012) and bacteria (Daudi et al. 2012).

lar space during plant defence responses. For instance, PLCPs such as tomato immune proteases Pip1 and Rcr3 increase in abundance upon biotic stress (Shabab et al. 2008; Ilyas et al. 2015). Taken together, diverse apoplastic proteins are involved in apoplastic plant immunity against an array of different pathogens. Some immune proteins increase in abundance during biotic stress and others do not. Several increasingly accumulating defence-related proteins are classified as PR proteins, while others are not. We thus propose an alternative grouping to describe defence-related, apoplastic proteomes. We suggest defining all proteins in the apoplast during biotic stress to be classified as part of a plant's immuno secretome. The immuno secretome thus comprises the apoplastic defence protein repertoire of a plant species and proteins can be secreted constitutively or in response to biotic stress. These proteins define the apoplastic proteomic landscape a pathogen has to navigate and overcome to colonise the apoplast. We consider proteins that are part of the immuno secretome to potentially be involved in plant immunity. In-

creasing abundance or activity upon biotic stress will further implicate them in immunity. A tool to study plant immuno secretomes is protein mass spectrometry (MS) of apoplastic fluid (AF) extracts (Martínez-González et al. 2018).

1.4 Co-evolution and positive selection in plant immunity

Antagonistic interactions place plants and pathogens under evolutionary pressure, selecting for individuals with the ability to detect, evade, damage or disarm mechanisms of defence and invasion of the other entity. This evolutionary pressure locks two species in an ongoing arms race which is not marked by an optimum final state, but defined by ongoing reciprocal adaptations and co-evolution (Dawkins and Krebs 1979). An example of co-evolving host and pathogen proteins is an apoplastic defence enzyme and a cognate pathogen-derived inhibitor. For the plant, co-evolution manifests itself in repeated selection of mutations that alter amino acids to reduce the interaction with the pathogen-derived inhibitor while maintaining enzymatic activity. For the pathogen, co-evolution selects for mutations resulting in amino acids that improve target inhibition. This co-evolution results in a disproportionate accumulation of mutations altering encoded amino acids (non-synonymous substitutions) compared to mutations not altering encoded amino acids (synonymous substitutions) which is defined as positive selection. Positive selection results in disproportionate accumulation of non-synonymous mutations in genes encoding co-evolving proteins and is thus an indicator for apoplastic plant proteins to be implicated in plant immunity.

To analyse positive selection, orthologous genes of closely-related species can be studied and their rates of non-synonymous over synonymous mutations compared (often referred to as dN/dS, Ka/Ks or ω). In plant immunity, molecular evolution was first studied for loci encoding single resistance genes (Thompson and Burdon 1992; McDowell et al. 1998; Caicedo et al. 1999; Stahl et al. 1999). Then, positive selection was found in a class I chitinase in the genus *Arabis*, demonstrating that secreted defence proteins are under positive selection as well (**Figure 2A**) (Bishop et al. 2000). Positive selection was next detected in apoplastic β-1,3-glucanases in the genus *Glycine* where residues under selec-

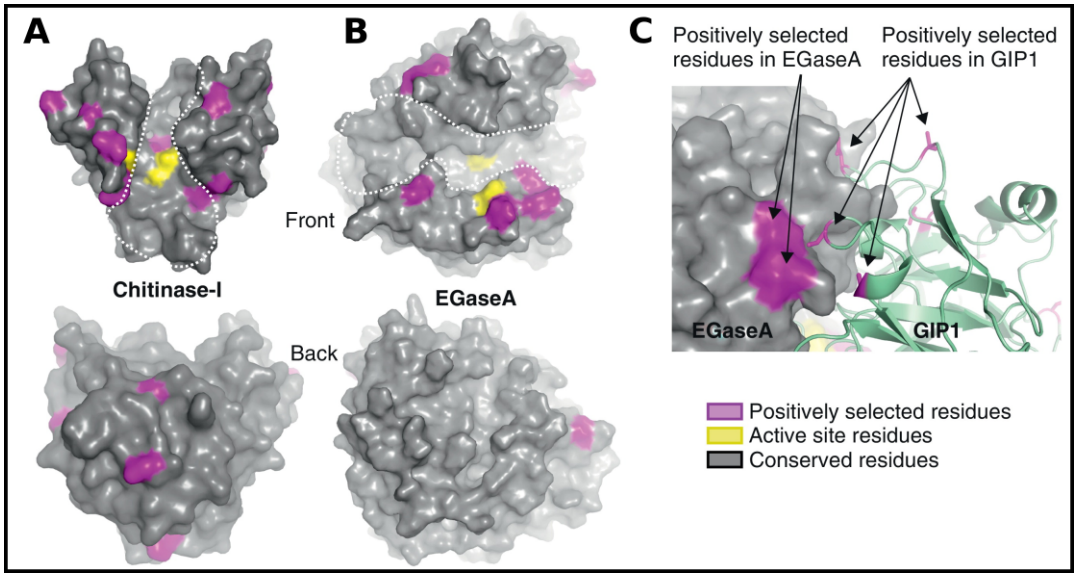


Figure 2: Positively selected residues in secreted defence proteins accumulate near active site residues. Amino acid residues encoded by codons under positive selection in class I chitinases in the genus *Arabis* (**A**) (Bishop et al. 2000) and *beta*-1,3-endoglucanase A (EGaseA) in the genus *Glycine* (**B**) (Bishop et al. 2005). (**C**) Residues encoded by codons under positive selection are in proximity to each other in the predicted enzyme (EGaseA) and cognate inhibitor (GIP1) complex (Damasceno et al. 2008). The Figure was taken and modified from Hörger and van der Hoorn 2013.

tion accumulated around the substrate binding groove, marking a 'ring of fire' (**Figure 2B**) (Bishop et al. 2005). In an inhibitor-enzyme model, these 'ring of fire' residues were in close proximity (<4.5 Å) to residues of the cognate *P.infestans* inhibitor GIP1 (**Figure 2C**) (Bishop et al. 2005). Shortly after, GIP1 was identified to be under positive selection in the *Phytophthora* genus, further corroborating co-evolution between *Glycine* β-1,3-glucanase and *Phytophthora* GIP1 (Damasceno et al. 2008). The PLCP C14 is under positive selection in wild potato species and residues under selection are in proximity to its cognate *P.infestans* inhibitor EpiC2B in a predicted PLCP-Cystatin complex (Kaschani et al. 2010; Kaschani and van der Hoorn 2011). Furthermore, tomato PLCPs Pip1 and Rcr3 and the subtilase P69B show increased accumulation of non-synonymous mutations in codons of surface-exposed residues surrounding the substrate binding groove (Shabab et al. 2008; Hörger and van der Hoorn 2013). These examples illustrate positive selection in apoplastic immune hydrolases potentially triggered by pathogen-derived inhibitors. Other drivers of positive selection, such as evasion of pathogen-derived proteases degrading immune hydrolases or adaptation to changing hydrolase substrates are

plausible (Buscaill and van der Hoorn 2021). Originally, positive selection was studied using sequences of orthologs generated by sequencing of PCR products. Necessity to sequence PCR amplicons was replaced by mapping genomic sequencing reads of closely-related species onto a reference genome to infer single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels) after whole genomes were sequenced and released. These include genomic sequencing data of crops and closely related wild species such as tomatoes (Consortium 2012; 100 Tomato Genome Sequencing Consortium 2014; Lin et al. 2014; Gao et al. 2019b) and soybean (Schmutz et al. 2010; Kim et al. 2010; Schmutz et al. 2010; Lam et al. 2010; Li et al. 2014; Liu et al. 2018). Wide-spread whole genome sequencing was enabled by drastically falling costs of genome sequencing (**Figure 3**). Reference-based variant calling has since shaped our understanding of domestication

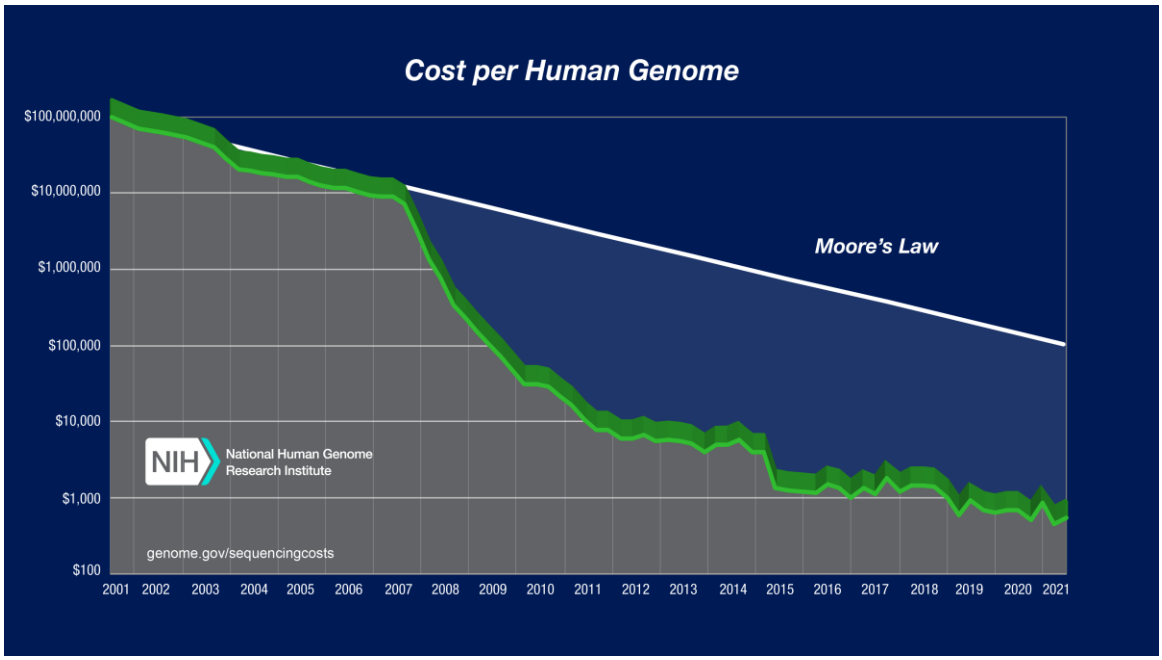


Figure 3: Reduction in sequencing cost measured in cost per human genome. The cost of sequencing a genome of the size of the human genome was reduced faster than Moore's law. Moore's law describes a long-term trend in compute power doubling every two years and, applied to DNA sequencing, the cost halving every two years. The figure was taken from the NIH.

history of different crops (100 Tomato Genome Sequencing Consortium 2014), but has not yet deepened our understanding of positive selection in immuno secretomes. In recent years, comparative genomics has moved from variant calling against a single reference

genome, to variant calling using pangenomes (Gao et al. 2019b; Liu et al. 2020; Qiao et al. 2021). Pangenomes capture the full breadth of genetic diversity of closely-related species as they entail all genomic regions of entailed genomes.

1.5 Structural prediction from co-evolution data

While breakthroughs in genome sequencing replaced sequencing of PCR amplicons to enable high-throughput analysis of positive selection, x-ray crystallography remained the gold-standard to analyse protein structures (Holcomb et al. 2017). 3D protein models based on x-ray cristallography-resolved structures were used to evaluate functional implications of codons under positive selection (Bishop et al. 2000; Bishop et al. 2005; Damasceno et al. 2008; Kaschani and van der Hoorn 2011). Residues were located on template-based 3D protein models to infer proximity to catalytic residues and establish the 'ring of fire' concept (Bishop et al. 2000; Bishop et al. 2005). Furthermore, putatively interacting residues under selection were described using template-based 3D models of enzyme-inhibitor complexes (Damasceno et al. 2008; Kaschani and van der Hoorn 2011). Replacing x-ray cristallography with accurate computational prediction of protein structures from the amino acid sequence was a long-standing limitation in biology. Advances in protein structure prediction had been assessed in the biannual critical assessment of methods of protein structure prediction (CASP) competition since 1994. In CASP14 in 2020, an advanced deep learning method called AlphaFold2 achieved accuracy scores comparable to experimentally resolved protein structures for the first time (**Figure 4**) (Jumper et al. 2021). AlphaFold2 infers protein structures from co-evolution encoded in deep amino acid multi sequence alignments (Jumper et al. 2021). While AlphaFold2 was trained on >100,000 experimentally resolved protein structures, it is now capable to accurately predict novel protein models without template structures of similar proteins (Jumper et al. 2021). Thus, low sequencing cost enables discovery of residues encoded by codons under positive selection in existing or newly-predicted protein structures.

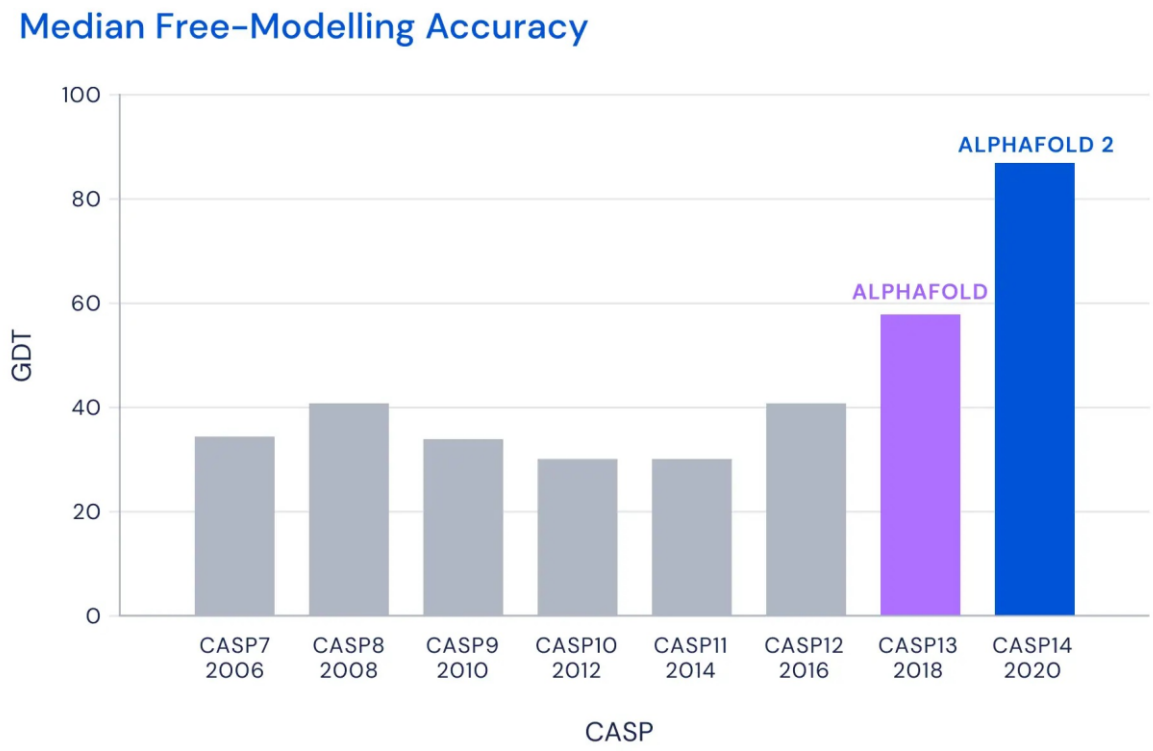


Figure 4: AlphaFold2 predicts high-quality protein models. The global distance test (GDT) ranges from 0-100 and measures the percentage of accurately predicted amino acids in a protein model. A GDT score of 90 of a predicted model is considered comparable to experimentally determined protein structures. The median GDT score of 92.4 means that the average error (RMSD) of AlphaFold2 predictions is 1.6 Å. The figure and data was taken from a Google Deepmind blog post.

1.6 Experimental implication in plant immunity

Both the increasing abundance upon biotic stress and positive selection implicate proteins in plant immunity. Further experimental validation is needed to proof involvement in plant immunity. The first step is the choice of pathosystem (plant and pathogen) as defence against fundamentally different pathogens such as bacteria, fungi or oomycetes involve different PR proteins, RLPs and RLKs (van Loon et al. 2006; Ngou et al. 2022) Only a few apoplastic immunity hubs, such as the PLCP Pip1, confer broad resistance to pathogens from a range of classes (Ilyas et al. 2015). Disease resistance can be divided into qualitative and quantitative traits conferring binary phenotypes and plants on a spectrum of resistance respectively (Corwin and Kliebenstein 2017). Qualitative disease

resistance is mediated by single loci and is amenable to experimental validation due to clear read outs. Quantitative disease resistance is encoded by multiple, small-effect genes. Confirming their involvement in plant immunity requires careful experimental design with ample replicates to be able to robustly distinguish between biological effects and technical noise (Corwin and Kliebenstein 2017). SAR-mediated increased disease resistance is quantitative (van loon 2000).

A.thaliana and wild tobacco (*Nicotiana benthamiana*) are common hosts to study plant immunity (Nishimura and Dangl 2010; Bally et al. 2018). Proteins are implicated in plant immunity by detecting induced physiological host defence responses such as ROS bursts, callose deposition or HR (Lamb and Dixon 1997; Balint-Kurti 2019; Wang et al. 2021b). While induction of defence responses are indicators for plant immunity, direct effects on pathogen growth is proof of involvement in disease resistance. Standard tools to quantify effects on pathogen growth include counting colony forming units, measuring bioluminescence and inferring pathogen biomass through PCR amplification or fluorophore-conjugated wheat germ agglutinin (WGA) (Fan et al. 2008; Ayliffe et al. 2013; Buscaill et al. 2021; Schnippenkoetter et al. 2021; Matsumoto et al. 2022). Furthermore, quantifying spread of bacteria involves bioluminescence and fluorescence (Misas-Villamil et al. 2013; Soldan et al. 2020) and symptom severity is based on predefined categories of disease onset (Planas-Marquès et al. 2018). Low throughput has been a bottleneck in testing putative quantitative resistance genes in disease assays. To increase throughput in plant-bacterial pathosystems, assays integrating bioluminescent bacteria have been developed recently (Soldan et al. 2020; Jutras et al. 2021; Matsumoto et al. 2022). Two *N.benthamiana* pathosystems have emerged to study qualitative and quantitative disease resistance. Agromonas, which is based on agrobacterium-mediated transient over expression and silencing of genes in combination with colony counts of *Pseudomonas syringae* (Buscaill et al. 2021). Whereas AgroLux is based on agrobacterium-mediated transient over expression and silencing of genes and resulting effects of plant physiological responses on agrobacteria, assessed by measuring bioluminescence of luxCDABE-expressing agrobacteria (Jutras et al. 2021). AgroLux can quickly and reliably assess HR

as a read out of qualitative disease resistance (Jutras et al. 2021).

1.7 Aims of the thesis:

Understanding plant immunity is important to breed resilient crops and sustainably feed a growing global population. The plant extracellular space, the apoplast, is a crucial battlefield in plant defence against attempted colonisation by pathogens. To understand apoplastic plant immunity, we need to know which proteins confer increased disease resistance. Therefore, identifying proteins in plant immunity by combining multiple lines of evidence helps build a foundation to understand plant immunity. Likewise, we need to understand fundamental similarities and differences between apoplastic immunity landscapes of different species as they provide the environments pathogens have to adapt to. The aims of this thesis are:

Chapter 2: The aim of this chapter was to identify tomato proteins implicated in apoplastic plant immunity by combining increased abundance upon biotic stress with analyses of positive selection. To study changes in abundance upon biotic stress, we analysed apoplastic fluids from three different Bion-treated tomato species with shotgun protein mass spectrometry. To study positive selection, we analysed locus-based orthologs of 12 tomato species by mapping genomic reads of wild tomato species against the *Solanum lycopersicum* reference genome. In this chapter, we also generated a reference proteome for *S.peruvianum* to annotate MS spectra. We implicated 21 proteins in plant immunity based on their increased abundance upon biotic stress and traces of positive selection. These putative tomato immune proteins include a protease inhibitor, a PR9 peroxidase and four P69 subtilases. P69 subtilases were the most abundant proteins in apoplastic fluids of biotic stress-induced tomato and displayed residues under positive selection surrounding the binding groove.

Chapter 3: The aim of this chapter was to establish fundamental differences and similarities between immuno secretomes of distantly-related eudicot species. We analysed apoplastic fluids from Bion-induced *Arabidopsis thaliana, Glycine max* and *Nicotiana benthamiana*. We discovered core differences and similarities between the secretomes

of six eudicot species using comparative proteomics: PR5 highly accumulated only in *A.thaliana*, P69 subtilases are only prominent in the tomato species while other subtilases were abundant in *N.benthamiana* and *G.max*. Similarities across the secretomes of the six species include the high accumulation of PR2 β-1,3-glucanases. This chapter highlighted different protein landscapes pathogens have to adapt to in different host species.

Chapter 4: The aim of this chapter was to establish a luminescence-based read out of the disease assay agromonas, called agromonasLux. We demonstrated that luminescence-based read out of luxCDABE-transformed *Pseudomonas syringae* pv. *tomato* DC3000 ΔhopQ (*Pto* DC3000) reflects enhanced plant immune responses in agroinfiltrated tissue transiently over expressing immunity-enhancing proteins. We could not identify enhanced resistance against *Pto* DC3000 when expressing three genes implicated in plant immunity from chapter 2.

Chapter 5: The aim of this chapter was to identify pathogen-derived inhibitors of six apoplastic tomato hydrolases using protein complex prediction by AlphaFold2 multimer. We established that AlphaFold2 multimer can accurately predict experimentally validated enzyme-inhibitor complexes when supplied with sufficient sequence information. By combining high-scoring, enzyme-inhibitor-resembling protein complexes with *in planta* expression data, we identified 13 novel, putative hydrolase inhibitors.

2 Tomato Immuno Secretomes

2.1 Introduction

Tomato is an important crop in modern agriculture and, with a global production of 183 billion tons, the vegetable with the largest production weight in the world (FAOstat). Protecting this widely-grown crop from losses to plant pathogens is thus of paramount significance. Wild tomato species have proven an invaluable resource for resistance breeding as disease resistance traits may have been lost during domestication (Schouten et al. 2019; Gao et al. 2019b). Resistance in wild tomatoes includes qualitative and quantitative traits providing protection against biotrophic and necrotrophic plant pathogens (ten Have et al. 2007; Finkers et al. 2008; Soltis et al. 2019; Kahlon et al. 2021). Recent research on tomatoes has focused on detecting novel qualitative resistance loci (Mesarich et al. 2018), understanding evolution and distribution of disease resistance networks (Stam et al. 2019a; Kahlon et al. 2020), further describing quantitative resistance (Soltis et al. 2019; Kahlon et al. 2021) and generating vast genomic resources for computational analyses (100 Tomato Genome Sequencing Consortium 2014; Lin et al. 2014; Gao et al. 2019b). While genomic studies and infection assays of wild tomatoes can be found manifold, little focus has been on proteomic adaptations in wild tomato upon biotic stress signalling, especially in the intercellular space, the apoplast. In plant immunity, plant-pathogen interactions in the apoplast can be decisive for the outcome of attempted infections (Doehlemann and Hemetsberger 2013). Reoccurring plant defence mechanisms in the apoplast involve the accumulation of highly abundant pathogenesis-related (PR) proteins and induced accumulation of lower-abundant proteins. Many of these accumulating proteins are proteases, glycosyl hydrolases and peroxidases (van Loon et al. 2006). Their functions in plant immunity seem to include the degradation of pathogen-derived components to damage structural integrity and release elicitors of immunity (Balasubramanian et al. 2012; Pusztahelyi 2018), activating host immunity-related components (Xia et al. 2004; Paulus et al. 2020), strengthening the host cell wall and contributing to production of antimicrobial phytoalexins and reactive oxygen species (Almagro et al. 2009). All to retard or stop

advancement of the pathogen infection. Plant pathogens attempt to evade and suppress host immunity components to establish a successful colonisation of the plant apoplast (Doehlemann and Hemetsberger 2013). Taken together, the apoplastic battlefield displays an intricate game of hide-and-seek-and-perturb in the struggle between hosts and pathogens. This ongoing arms race causes selection pressure on antagonistic plant- and cognate pathogen-derived proteins that manifests in a genetic footprint of positive selection (Bishop et al. 2000; Bishop et al. 2005; Damasceno et al. 2008; Shabab et al. 2008; Kaschani et al. 2010; Kaschani and van der Hoorn 2011). Combining the breadth of pathogen resistance described in genetically diverse wild tomato species and the importance of the apoplastic battlefield in plant immunity we asked ourselves two questions: First, what are differences and similarities in how different tomato species defend their apoplasts during biotic stress? Second, can we identify novel apoplastic tomato proteins implicated in apoplastic immunity? To address these questions, we will first induce a generic biotic stress responses in a broad range of tomato species and extract their apoplastic fluid. Biotic stress will be induced by treating plants with Bion, a commercial pesticide containing 50% w/w BTH and undisclosed additives, which is used in agriculture to activate "the plant's natural defense mechanisms" (Syngenta's Bion). BTH is a functional analogue of the plant biotic stress hormone salicylic acid (Bektas and Eulgem 2015). Second, we will generate a proteome for *S.peruvianum* from genomic reads to be able to analyse apoplastic fluid extracts via mass spectrometry. Third, we will compare adaptations in apoplastic proteomes of three different tomato species (*S.lycopersicum, S.peruvianum* and *S.pennellii*) induced with BTH to establish their differences and similarities. Third, we will analyse molecular evolution in the form of positive selection to further implicate biotic stress-induced apoplastic tomato proteins in plant immunity.

2.2 Materials and Methods

2.2.1 Apoplastic Fluid extraction

To extract apoplastic fluid, we syringe-infiltrated leaflets of 3-4 week old plants with 100 mM phosphate buffer (pH 6.5) using a disposable 50 ml syringe. The choice of phos-

phate buffer was based on its comparably low level of extracted cytoplasmic contaminants (Witzel et al. 2011). After infiltrating leaflets, we dried their surfaces and rolled them up carefully to fit into a disposable 20 ml syringe. Each syringe was placed into a sterile 50 ml falcon tube and stored on ice until all samples were ready for centrifugation. Next, we extracted apoplastic fluid from infiltrated leaflets by centrifugation for 25 minutes, at 4°C and 1000 x g for *S.lycopersicum* and *S.pennellii* and 2000 x g for *S.peruvianum*. Extraction force for *S.peruvianum* had been higher as preliminary data had suggested improved apoplastic fluid extraction at higher force compared to *S.lycopersicum* and *S.pennellii*. After extraction, apoplastic fluid was aliquoted and flash frozen in liquid nitrogen as quickly as possible. One biological replicate was comprised of leaflets from different leaves. Leaflets had been weighed before syringe-infiltration and the amount of extracted apoplastic fluid measured to later calculate the extracted amount per fresh weight.

Total extracts were made from weighed-in leaf discs taken from the same plants used for apoplastic fluid extraction. Leaf discs were flash frozen in liquid nitrogen in 1.5 ml Eppendorf tubes with 2 metal beads. Leaf tissue was pulverised with a tissue lyser at a frequency of 30 per seconds for 1 minute. The tissue lyser blocks had been pre-cooled at −80°C beforehand. Pulverised tissue was mixed with 200 µl of ice-cold PBS buffer (0.137 M NaCl, 0.0027 M KCl, 0.01 M Na_2HPO_4 and 0.0018 M KH_2PO_4) on ice and leaf debris removed by centrifugation at 13,000 x g and 4°C for ten minutes. Supernatant was separated from the leaf debris and stored on ice to be used directly or flash frozen in liquid nitrogen.

2.2.2 Tomato genomic resources used

In this chapter, we used publicly available genomic sequencing reads from twelve different tomato species (**Table 2**).

Table 2: Genomic reads of twelve tomato species

Species	Accession	*SRAID*[1]	Publication
S.lycopersicum	Heinz-1706	SRR1572628	(Lin et al. 2014)
S.lyc. var. *cer.*[2]	BGV006865	SRR7279628	Bioproject PRJNA454805
S.pimpinellifolium	LA2093	SRR12039813	Wang et al. 2020b
S.cheesmaniae	LA0483	ERR418087	100 Tomato Genome Sequencing Consortium 2014
S.arcanum	LA2157	ERR418092	100 Tomato Genome Sequencing Consortium 2014
S.neorickii	LA2133	ERR418090	100 Tomato Genome Sequencing Consortium 2014
S.huaylasense	LA1983	ERR418095	100 Tomato Genome Sequencing Consortium 2014
S.chilense	LA3111	SRR13259416	Stam et al. 2019b
S.corneliomuelleri	LA0118	ERR418061	(100 Tomato Genome Sequencing Consortium 2014)
S.peruvianum	LA1954	ERR418094	100 Tomato Genome Sequencing Consortium 2014
S.habrochaites	LYC4	ERR410237	100 Tomato Genome Sequencing Consortium 2014
S.pennellii	LA0716	ERR418107	100 Tomato Genome Sequencing Consortium 2014

[1] SRA ID: NCBI's Sequence Read Archive' Sequencing Run ID. [2] *S.lyc.* var. *cer.*: *Solanum lycopersicum* var. *cerasiforme*.

2.2.3 Estimating genetic distances between tomato species

In the absence of genome assemblies and proteomes for all twelve tomato species, we estimated genetic distances between tomato species using kmer-based distance matrices. We first downloaded their genomic sequencing reads described in **Table 2** using NCBI's SRAtoolkit (Leinonen et al. 2011). Genomic sequencing reads were quality trimmed using trimmomatic v0.39 with the following settings 'LEADING:3 TRAILING:3 SLIDINGWINDOW4:15 MINLEN:36' (Bolger et al. 2014b). Then, we interleaved paired sequencing reads using BBMap's reformat.sh script (Bushnell 2014). Next, we computed mash distances to approximate whole genome distances between species (Ondov et al. 2016). Mash distances rely on reducing the complexity of whole genomes to kmers that are converted to hashes, integers. Next, a subset of these hashes can be chosen to define a sketch. Sketches from different genomic sequencing reads were then compared to each other to calculate how many hashed kmers they have in common which can be used to approximate distances between the whole genome sequencing sets the sketches were derived from. The whole-genome distance matrix was then used to generate a neighbor-joining phylogenetic tree. The entire workflow is implemented in mashtree 1.2.0. We used a sketch size of 100,000 and set the minimum depth of kmers to 5. The phylogenetic tree was visualised using iTOL (Letunic and Bork 2021).

2.2.4 Enzymatic characterisation of apoplastic fluid extracts

Protein concentrations of apoplastic fluid and total extract samples were measured with DC™ Protein Assay Kit and all procedures were based on the BioRad's manual. All measurements were taken in transparent 96-well plates. First, working reagent A' was prepared in advance by mixing 20 µl of reagent S with one ml of reagent A. For a reference standard, we used BSA protein standard at 0.125, 0.75 and 1.5 mg/ml concentrations. In wells of a 96-well plate, we added 5 µl of reference BSA standard solution or protein sample to measure in technical triplicates, then we added 25 µl of reagent A' and 200 µl reagent B. Next, samples were incubated in the dark for 15 minutes. Finally, we measured absorbance at 750 nm. Data was analysed in a python jupyter notebook (Kluyver et al. 2016; Homma 2022d). Cytoplasmic contamination of apoplastic fluid samples was calculated by comparing enzyme activity of malate dehydrogenase (MDH) per gram fresh weight in apoplastic fluid samples to total extract samples. This approach was based on MDH activity determined in Grosse-Holz et al. 2018. Measurements were taken in 96-well plates and the total reaction volume was 200 µl. This defined the pathlength of 0.51 cm. The reaction took place in Tris-HCl buffer (50 mM, pH7.5) with the substrates oxaloacetic acid (0.8 mM) and NADH (0.4 mM). Apoplastic fluid was diluted 1:10 and total extract samples were diluted 1:1000. All apoplastic fluid samples were added to wells before adding reagents in Tris-HCl buffer with a multi-channel pipette in the plate reader to reduce bias from delayed starts in enzymatic reactions. Measurements were taken in intervals of 15-30 minutes and average volume of extracted apoplastic fluid was based on a previously determined average per g of fresh leaf. MDH activity of apoplastic fluids was measured for the same apoplastic fluid conditions used to generate biological material for later mass spectrometry analysis. For *S.pennellii*, the weight of four leaf discs had not been measured which is why the average of the same number of leaf discs of *S.lycopersicum* and *S.peruvianum* was used to infer the weight of *S.pennellii* leaf discs.

2.2.5 Activity-based protein profiling (ABPP) of serine hydrolases

A total reaction volume of 35 µl contained 30 µl of apoplastic fluid, Tris-HCl buffer (pH 6.8, final concentration 50 mM) and FP-TAMRA probe (final concentration 0.2 µM) dissolved in DMSO. The reaction mix was incubated in the dark for 60 minutes. Afterwards, adding 7 µl of gel loading buffer (200 mM Tris-HCl, pH 6.8, 400 mM DTT, 8% SDS, 0.4% bromphenol blue, 40% glycerol) with subsequent incubation at 95°C for 5 minutes stopped the reaction. FP-labelled samples were separated via SDS PAGE. The fluorescently-labelled proteins were detected with a Typhoon scanner and intensities quantified using imageJ (Schneider et al. 2012).

2.2.6 Generating a *Solanum peruvianum* reference proteome

We downloaded genomic sequencing reads of *S.peruvianum* LA1954 from NCBI's sequence read archive of the experimental accesion ERX384394 and sequencing run ERR418094. Sequencing reads were quality trimmed using sickle 1.21. Quality-trimmed reads were mapped against the GenBank GCA_000188115.3 SL3.0 reference assembly from NCBI using BWA-MEM v0.7.17 with a maximum insert size of 750 bp (Li 2013). Then, we sorted mapped reads and called single nucleotide variants and small inserts and deletions using Samtools v1.7 and filtered to only retain variants with a phred score >40 (Li et al. 2009; Li 2011). Then, we *in silico* mutated the *S.lycopersicum* GCA_000188115.3 into the *S.peruvianum* LA1954 genome by incorporating all high-quality variants using a custom-made python script named vcf_to_genome_tomato.py deposited (Homma 2022d). For gene annotation of the *S.peruvianum* LA1954 quasi-reference genome based on homology of known tomato proteins using scipio 1.4.1 (Keller et al. 2008), we used a non-redundant pan-proteome of published tomato proteomes from *S.lycopersicum* GCF_000188115.4, *S.pennellii* GCF_001406875.1, *S.chilense* GCA_006013705.1 and S.pimpi-nellifolium ps .pimp.maker.proteins.FINAL.faa. The tomato pan proteome was generated using CD-HIT v4.7 with a sequence identity threshold of 0.95 (Li and Durbin 2009). Scipio-based gene annotations did not correct for premature stop codons in *S.peruvianum* compared to other tomato homologs. We thus removed all gene models that contained a stop codon be-

fore less than 70% of their predicted total length using a custom python script. A similar approach to Scipio-based gene annotation of homologs had been taken in Kourelis et al. 2019. BUSCO scores were determined with BUSCO v5.1.3 for the solanales_odb10.2020-08-05.tar.gz data set as a reference (Simão et al. 2015).

2.2.7 Preparing protein samples for mass spectrometry analysis

Samples were prepared using a filter aided sample preparation (FASP) protocol established in the lab by Dr Kyoko Morimoto. As filters, we used Vivacon 500 (10,000 MWCO) and for each sample, we prepared a total amount of 50 µg of protein. Filters were prepared by adding 200 µl of 0.1% trifluoroacetic acid (TFA) in 50% acetonitrile and centrifuged at 15,000 x g for 15 minutes at room temperature (RT). Next, 200 µl of 8 M urea in 100 mM triethylammonium bicarbonate buffer (TEAB) were added with 50 µg of protein sample and centrifuged at 15,000 x g for 15 minutes followed by four washing steps with 8 M urea in 100 mM TEAB. Next, proteins were denatured by again adding 8 M urea in 100 mM TEAB in addition to 4 µl of 500 mM (tris 2-carboxyethyl) phosphine Hydrochloride (TCEP HCl), 20 µl of 500 mM 2-chloroacetamide, incubated in the dark at RT for 30 minutes and centrifuged at 15,000 x g for 15 minutes at RT. The filter was washed twice with 200 µl 6 M urea in 50 mM TEAB and centrifuged at 15,000 x g for 15 minutes at RT. Next, lys-C solution was added at a ratio of 5 µg of lys-C per 200 µg of initially applied protein. Lys-C had been dissolved in 1 M urea in 50 mM TEAB at a concentration of 5 µg/100µl. After over-night incubation at 37°C, we added trypsin at a ratio of again 5 µg trypsin per 200 µg of initially applied protein and incubated for 6-8 hours at 37°C. Trypsin had been dissolved in 100 µl 50 mM TEAB at a concentration of 5 µg/100µl. Finally, filters were centrifuged at 15,000 x g for 15 minutes at RT and washed with 150 µl 0.1% TFA, then 150 µl 0.1% TFA in 50% Acetonitrile and the flowthrough stored in a protein low-binding tube. Samples were not dried but sent on dry ice to Dr. Farnusch Kaschani at the Analytics Core Facility Essen, Germany.

2.2.8 Mass spectrometry data processing and statistical analyses

Protein mass spectrometry data was analysed by Dr Farnusch Kaschani. Peptides were identified using default settings in MaxQuant v1.6.10.43 using Andromeda as a search engine with label free quantification (LFQ) and match between runs turned on (Cox and Mann 2008; Cox et al. 2011; Cox et al. 2014). LFQ values inferred by MaxQuant are based on MaxLFQ normalisation and all referenced LFQ values in this chapter were based on MaxLFQ normalisation (Cox et al. 2014). Static modification included carbamidomethyl and dynamic modifications included N-terminal acetylation, oxidation and deamidation. Oxidation was a modification included in quantification. Searches were performed against the *Solanum lycopersicum* reference proteome ITAG4.0, *Solanum pennellii* RefSeq proteome GCF_001406875.1 and our own *Solanum peruvianum* reference proteome (**section 2.2.6**) and a database of known contaminants. To estimate false positive rates, MaxQuant furthermore generates a decoy database by reversing and processing the reference proteome which generates nonsensical peptides with preserved overall amino acid composition (Elias and Gygi 2007; Käll et al. 2008). To benefit from identifications from total leaf extract protein samples in identifying peptides in apoplastic fluid samples and to not distort values during the normalisation step, we analysed apoplastic fluid and total extract samples as separate parameter groups with default minimal label-free quantification (LFQ) ratio count of 2.

2.2.9 Analysis of proteomics data of tomato apoplastic fluids

All filtering steps were the same for each species with their cognate predicted secretomes and analysis were done in Jupyter-Notebooks (Kluyver et al. 2016; Homma 2022d). Since tryptic digest of different proteins can result in identical peptides, MS analyses can not always distinguish between different proteins. Groups of proteins that match identified peptides and for which insufficient evidence was collected to confidently identify the individual proteins are thus grouped into protein groups. These protein groups mostly contain proteins with high sequence similarity. In MaxQuant analyses, protein groups are the smallest unit of confident identification of proteins and many protein groups only

contain a single protein. First, protein groups were filtered by removing groups identified as contaminants, identified by matching only and labelled as reverse. Reverse proteins derived from the decoy search database (reversed reference proteome) are false positives. Next, only protein groups identified in at least three out of four replicates in either or both treatments were considered to be detected robustly. No values were imputed. PCA analyses were done with centred protein groups implemented in sklearn (Pedregosa et al. 2011). Log2-transformed LFQ intensities were centred per protein group by subtracting the protein group's mean from each sample. Orthogroups were generated for whole proteomes of *S.lycopersicum*, *S.pennellii*, *S.peruvianum* and *N.benthamiana* using Orthofinder 2.0 (Emms and Kelly 2019). Intensities for each orthogroup and species were determined by the sum of untransformed LFQ intensities of each contributing protein. Log2-transformed, summed-up LFQ intensities were centred per orthogroup and species by subtracting the orthogroup's and species' mean from each respective sample. To reduce correction for multiple testing, we selected a minimal log2FC threshold of 0.9 for proteins to be considered for statistical analysis of differential abundance. This value was based on a preliminary MS data exploration of known tomato PR7, P69B, which visibly accumulates on protein gels and is known to increase in abundance upon Bion treatment (Tian et al. 2004). Changes were calculated using unpaired, two-sided pairwise t-tests. For each species, we corrected p-values of all pairwise comparisons for multiple testing using Benjamini-Hochberg procedure. We used a significance threshold of 0.05 for multiple testing corrected p-values (sometimes referred to as q-values). To derive a proxy for protein concentration, we calculated percentage shares of total LFQ intensities of orthogroups by dividing the untransformed LFQ intensity of an orthogroup by the sum of untransformed LFQ intensities of the respective sample. Functional annotations of proteins were determined using InterPro Scan v5.50.84.0 (Jones et al. 2014). Functional annotation of orthogroups was based on Pfam annotations of *S.lycopersicum* homologs. If *S.lycopersicum* did not encode a protein for an orthogroup, Pfam annotations of either *S.pennellii* or *S.peruvianum* defined the functional annotation of the orthogroup.

2.2.10 Ploidy analysis

To estimate the ploidy of the twelve tomato species, we compared heterozygous kmer coverages of quality-trimmed whole-genome sequencing data using smudgeplot v0.2.3dev (Ranallo-Benavidez et al. 2020). The same quality-trimmed genomic sequences used to generate whole genome species trees were used here (**section 2.2.3**).

2.2.11 Positive selection analysis

The positive selection workflow was divided into four distinct pipelines. In the variant calling pipeline (Homma 2022g) quality-trimmed reads from whole genome species trees were reused and FastQC integrated to assess read quality parameters (**section 2.2.3**). We mapped reads against the Sol4.0 *S.lycopersicum* reference genome assembly using BWA-MEM v0.7.17 (Li 2013). Mapped reads were processed and sorted using Samtools v1.7 (Li et al. 2009; Li 2011). InDels were realigned using GATK v3.8-1-0-gf15c1c3ef (Auwera and O'Connor 2020). Variants were called using bcftools v1.7 using a phred score of 20 as a cut off (Li 2011). In the haplotype phasing (Homma 2022e) workflow, we phased reads using whatshap v1.0 (Martin et al. 2016). In the allele generation pipeline (Homma 2022c), we identified coding sequences from loci including wild tomato variants using exonerate v2.4.0 (Slater and Birney 2005). In the positive selection pipeline (Homma 2022f), we generated protein multi sequence alignments using mafft v7.475 and created codon-guided nucleotide multi sequence alignments using pal2nal v14 (Suyama et al. 2006; Katoh and Standley 2013). Maximum likelihood phylogenetic gene trees were generated using RAxMLHPC v8.2.12 (Stamatakis 2014). These codon-guided multi sequence alignments in combination with maximum likelihood guide trees were used to analyse site-wise positive selection by phylogenetic analysis by maximum likelihood implemented in codeml which is part of the PAML v4.9 package (Nielsen and Yang 1998; Yang et al. 2000; Yang et al. 2005). Positive selection was determined by conducting likelihood ratio tests for nested gamma models M7 and M8 (Yang et al. 2000). We corrected for multiple testing using the Benjamini-Hochberg procedure and selected a significance threshold of 0.05. To identify codons under positive selection, we set the bayes empricial bayes posterior probability threshold

to 0.8 to retain a high true positive rate while maintaining a low false negative rate (Yang et al. 2005). To compare significance levels to genome-wide positive selection analysis in strawberry (Qiao et al. 2021), we added 19,084 simulated p-values following a normal distribution with a mean of 0.5, standard deviation of 0.2 and 0 and 1 as lower and upper limits to p-values of our 51 genes analysed for positive selection.

2.3 Results

2.3.1 Apoplastic fluids from different tomato species

To study apoplastic landscapes of distantly-related tomato species, we grew and extracted AF from *S.lycopersicum, pimpinellifolium, chilense, peruvianum* and *pennellii*. This selection covered three of four tomato species clades and included tomato species with published reference genomes and proteomes for *S.lycopersicum*, *pimpinellifolium, chilense* and *pennellii* (Consortium 2012; Bolger et al. 2014a; Razali et al. 2018; Stam et al. 2019b). To estimate genomic relationship between different tomato species, we generated a whole-genome sequencing reads-based phylogenetic tree for 12 different tomato species using Mashtree (Katz et al. 2019). First, we observed overall high bootstrap support for different clades. Our results are in line with the distinct species clades *Lycopersicon* and *Arcanum* (Peralta et al. 2008) (**Figure 5**). To induce a biotic stress

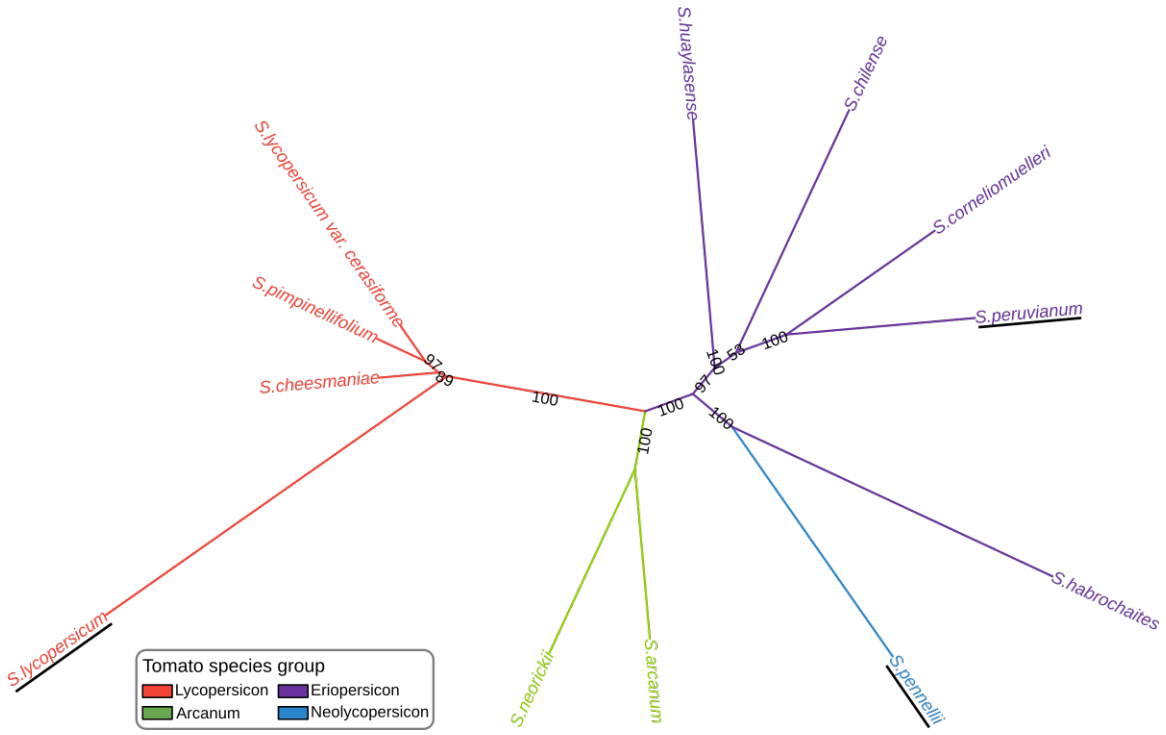


Figure 5: Unrooted, kmer-based phylogenetic tree of twelve tomato species. Clades are coloured based on originally proposed tomato species groups (Peralta et al. 2008). Species continued for comparative proteomics were underlined. Neighbor-joining tree inferred from whole-genome distance matrix based on Mash distances. Sketches contained 100,000 min-hashed kmers and kmers were sampled from whole-genome Illumina sequencing reads. All bootstrap values are depicted.

response, clonally propagated cuttings from the same individual were watered for three days, each time with 10 ml water or with 10 ml of 1 mM of the functional salicylic acid analogue acibenzolar-s-methyl (BTH). Syngenta's Bion contains 50% (w/w) BTH and was used for biotic stress induction. Six days after the first Bion treatment, AF was extracted. We observed Bion-induced accumulation of pathogenesis-related (PR) proteins in AF from *S.lycopersicum, peruvianum* and *pennellii* (**Figure 6A**). *S.pennellii* samples

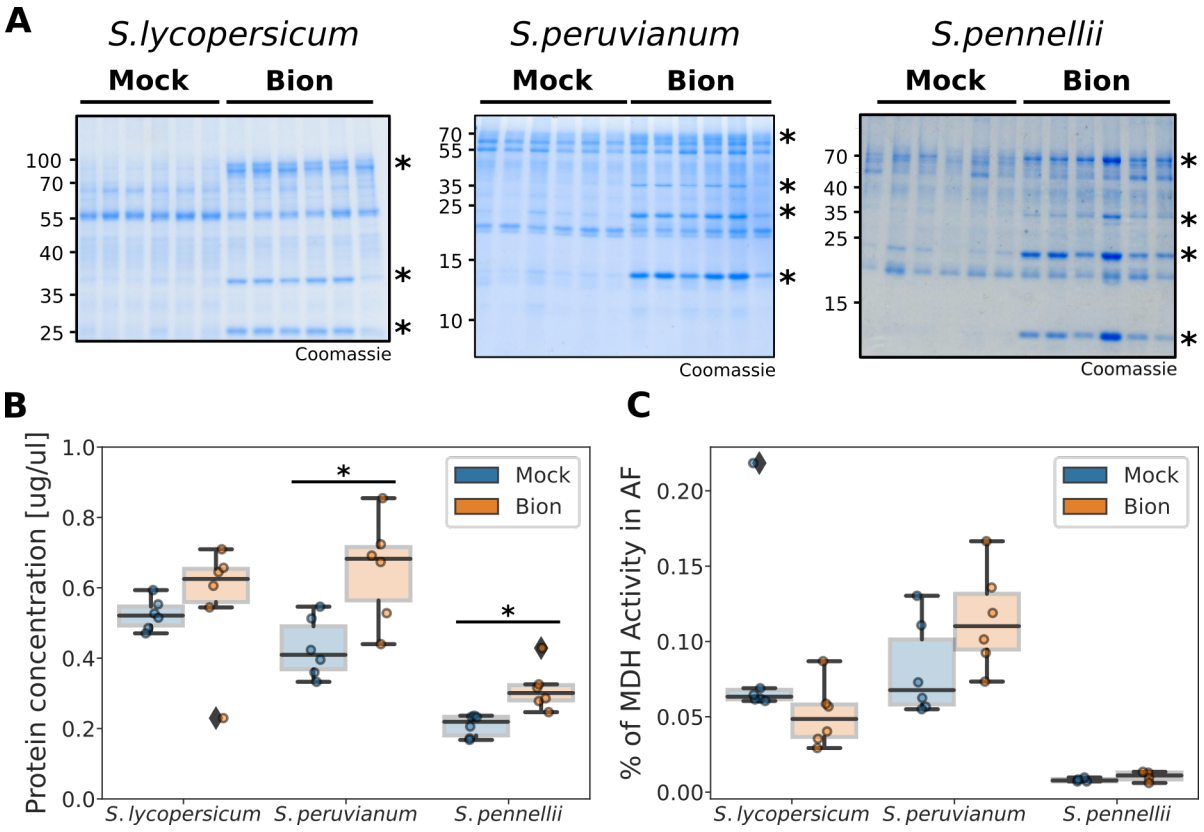


Figure 6: Bion-induced protein accumulation in apoplastic fluid of tomato species. (**A, B, C**) All apoplastic fluid samples represent separate biological replicates from the same experiments. (**A, B**) Apoplastic fluid extracts from gels and enzymatic assays are aliquots of the same samples. (**A**) Asterisks indicate accumulating pathogenesis related (PR) proteins. For *S.lycopersicum*, bands < 25 kDa had not been resolved. (**B**) Protein concentration of AFs from unstressed (Mock) or biotic stress-induced (Bion) tomato plants. (**C**) Cytoplasmic contamination of AFs was assessed expressing MDH activity in AFs relative to MDH activity in total extracts. (**B, C**) Centre lines are the median, bounds of boxes the 25th and 75th percentiles and whiskers 1.5 x interquartile range (IQR) from 25th and 75th percentiles. Diamonds represent outliers. Benjamini-Hochberg, false discovery rate-corrected p-values of two-sided, unpaired t-tests are indicated as: $* < 0.05$.

showed lower protein concentration than samples from *S.lycopersicum* and *S.peruvianum* (**Figure 6B**). Bion treatment increased the protein concentration in AF by 52.2% and

50.7% for *S.peruvianum* and *S.pennellii* respectively. We observed an insignificant Bion-induced increase in protein concentration of only 7.8% in *S.lycopersicum* (FDR-corrected p-value of two-sided, unpaired t-test: 0.588) (**Figure 6 B**).

To measure cytoplasmic contamination, we determined activity of the intracellular enzyme malate dehydrogenase (MDH) in AF and total extracts of leaf discs that had been taken from the same batch of plants. First, all AFs showed low MDH activity, ranging from 0.006% to 0.22% of that found in total extracts. To put cytoplasmic contamination into context, published relative MDH activity in AF extracts of the Solanaceae species *N.benthamiana* were in a similar range between approximately 0.05% and 0.2% (Grosse-Holz et al. 2018). Second, AFs from *S.pennellii* had relatively low MDH activity compared to AF from *S.lycopersicum* and *S.peruvianum* (**Figure 6C**). Third, within-species comparison showed high variability in relative MDH activity and no comparison between mock- and Bion-induced AFs showed significant differences (FDR-corrected p-values of two-sided, unpaired t-tests had a p-value of 0.246 for all three within-species comparisons). Therefore, Bion-induced increase in protein concentration in AFs of (wild) tomato were not caused by increased cytoplasmic contamination.

Next, we investigated changes in active serine hydrolases in AFs using the activity-based probe FP-TAMRA. FP-TAMRA is a fluorophosphonate (FP) conjugated with tetramethyl rhodamine (TAMRA) fluorophore (Liu et al. 1999; Patricelli et al. 2001). FP reacts with active site serine residues causing fluorescent labelling of active serine hydrolases, including P69 subtilases which clearly accumulate as PR proteins in coomassie-stained SDS-PAGE protein gels (**Figure 7**). To quantify active P69-like subtilases, we measured intensities of fluorescent bands using ImageJ and standardised the intensities for each species relative to the mean intensity detected for respective mock-induced AFs (**Figure 7**). For *S.lycopersicum, S.peruvianum* and *S.pennellii*, P69 activity increases by 239%, 98% and 222% respectively (respective FDR-corrected p-values of one-sided, unpaired t-tests: 0.00000023, 0.0048 and 0.00079). We thus showed a strong increase in activity of P69-like subtilases upon biotic stress treatment in distantly related tomato species.

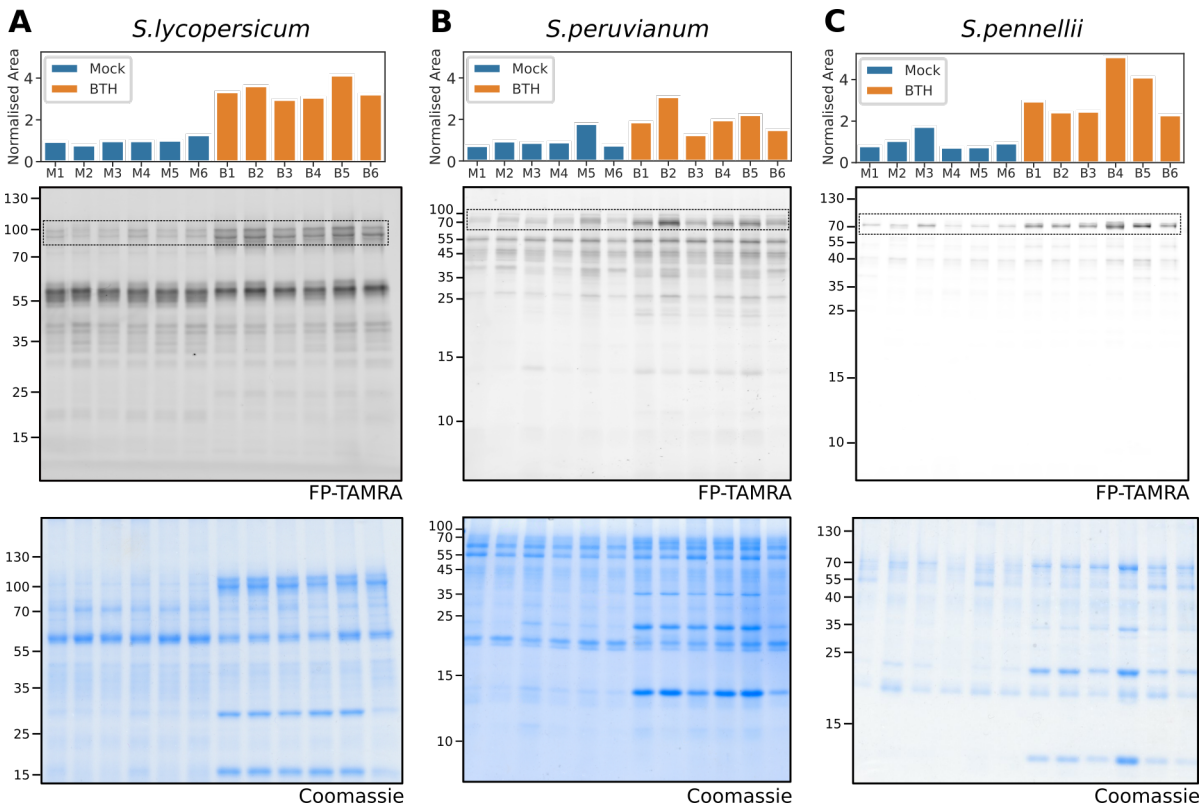


Figure 7: Apoplastic subtilase activity increases upon biotic stress in three different tomato species. (**A, B, C**) Apoplastic fluid (AF) extracts had been separated via SDS-PAGE on 12% (**A**), 18% (**B**) and 15% (**C**) gels. Proteins gels are identical to the gels shown in **Figure 6 A**. Bars above scanned gels indicate the intensity of bands measured for P69 subtilases. Boxes of dashed lines indicate the areas for which band intensities were quantified using ImageJ. Values for each bar were normalised relative to the mean intensity detected in respective mock-treated AF extracts.

2.3.2 Generating a *Solanum peruvianum* reference proteome

To generate a reference proteome for *S.peruvianum*, we downloaded and quality-trimmed genomic sequencing reads of *S.peruvianum* LA1954 from a large tomato genome sequencing project (100 Tomato Genome Sequencing Consortium 2014). The *S.peruvianum* genomic reads had been generated from a plant of the same LA1954 accession which was used for Bion treatment and AF extraction. First, we mapped quality-trimmed reads against the SL3.0 *S.lycopersicum* GenBank reference genome assembly and called variants between the mapped genomic sequencing reads and the reference genome (Consortium 2012; 100 Tomato Genome Sequencing Consortium 2014). Then, we *in silico* mutated the reference genome to resemble the *S.peruvianum* genome based on high-quality SNPs and small indels. Thereafter, we used a non-redundant tomato pan-proteome of pub-

lished tomato proteomes of *S.lycopersicum, S.pimpinellifolium, S.chilense* and *S.pennellii* to structurally annotate the *S.peruvianum* genome based on homology (Consortium 2012; Bolger et al. 2014a; Razali et al. 2018; Stam et al. 2019b). The workflow is conceptually explained in **Figure 8**. For a more detailed explanation of the data sets and tools used, see Materials and Methods **section 2.2.6**. After generating the proteome, we benchmarked it against published tomato proteomes of *S.lycopersicum, S.pimpinellifolium, S.chilense* and *S.pennellii* for size and completeness. The new *S.peruvianum* LA1954 proteome con-

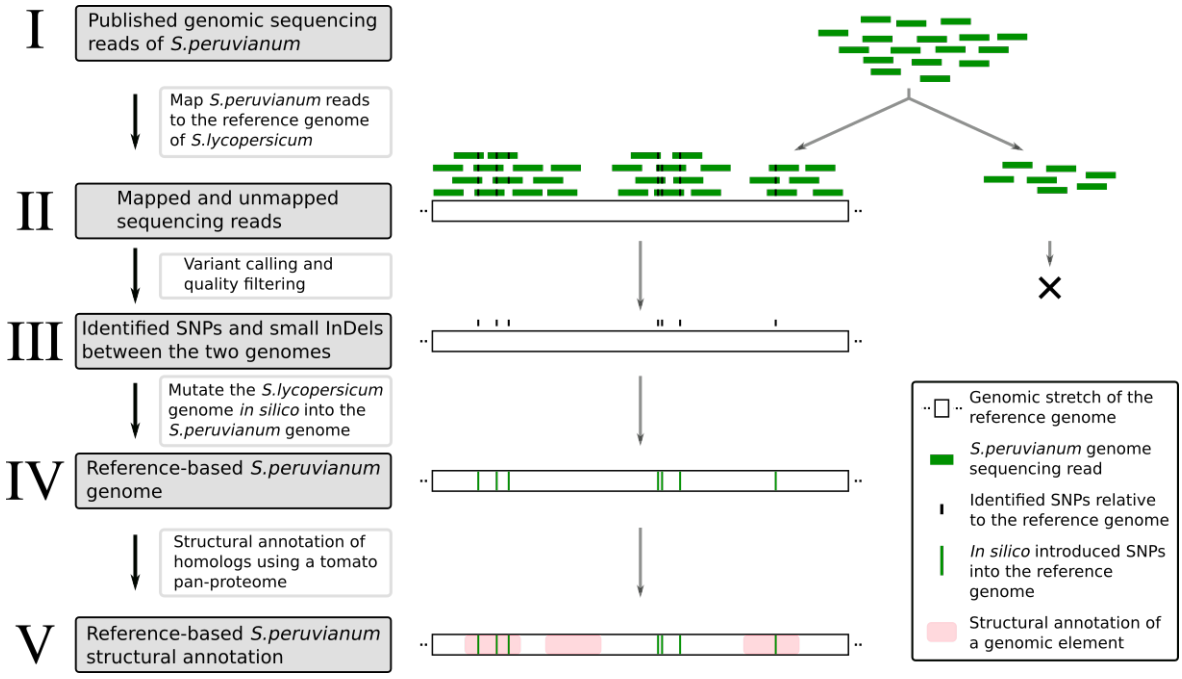


Figure 8: Workflow to generate a reference-based *S.peruvianum* proteome. (**I**) Published genomic sequencing reads from *S.peruvianum* were downloaded and quality trimmed. (**II**) Quality-trimmed genomic sequencing reads from *S.peruvianum* were mapped onto the Sl3.0 tomato reference genome. Unmapped sequencing reads were discarded. (**III**) Robust variants such as SNPs and small InDels were called relative to the *S.lycopersicum* reference genome. (**IV**) Robust variants replaced reference nucleotides in the *S.lycopersicum* genome to generate a quasi reference genome of *S.peruvianum*. (**V**) A non-redundant tomato pan-proteome based on published proteomes of four tomato species was used to structurally annotate the quasi *S.peruvianum* reference genome based on homology.

tains 52,929 proteins and is larger than the four other published tomato proteomes which ranged from 25,744 for *S.pimpinellifolium* to 38,023 for *S.pennellii* (**Figure 9A**). Next, we analysed Benchmarking Universal Single-Copy Orthologs (BUSCO) scores to assess completeness of the proteomes based on a Solanales reference databases. Completeness of a proteome was measured by combining the fractions of detected single-copy and duplicated

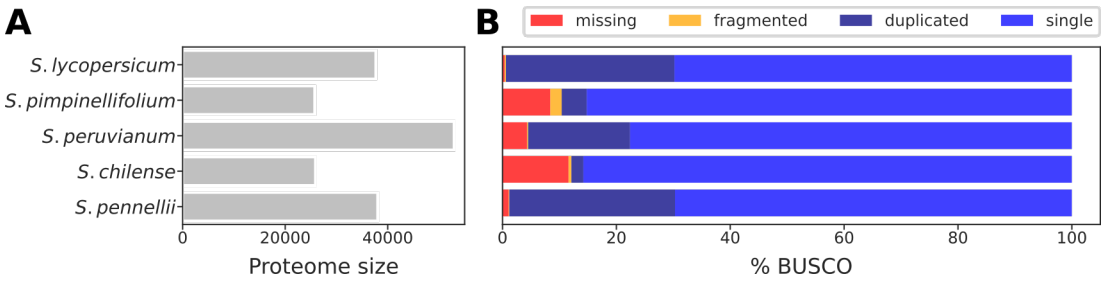


Figure 9: The large *S.peruvianum* reference proteome is as complete as published tomato proteomes. (**A**) Comparing proteome sizes of published tomato proteomes. (**B**) Analysing different tomato proteomes based on completeness of Solanales BUSCO searches.

Solanales BUSCO proteins. The *S.peruvianum* proteome showed the third most complete set of Solanales BUSCO proteins after *S.lycopersicum* and *S.pennellii* (**Figure 9B**). The *S.peruvianum* proteome also contained the third lowest amount of duplicated Solanales BUSCO proteins, after *S.chilense* and *S.pimpinellifolium*. Furthermore, the *S.peruvianum* proteome tied *S.lycopersicum* for second fewest fragmented Solanales BUSCO proteins, just after *S.pennellii*. Last, our newly generated proteome contained third fewest missing Solanales BUSCO proteins, after *S.lycopersicum* and *S.pennellii*. Taken together, our proteome of *S.peruvianum* is likely too large, but shows similiar completeness to published proteomes of other tomato species.

2.3.3 Immuno secretomes of different tomato species

To understand similarities and differences between biotic stress-induced apoplastic proteomic landscapes of distantly-related tomato species, we first generated protein mass spectrometry (MS) data of AF from mock- and Bion-treated *S.lycopersicum, S.peruvianum* and *S.pennellii* comprising four biological replicates each. Upon visual inspection of all log2-transformed intensities of protein groups, values appeared normally distributed and did not contain outlier samples for any treatment within any of the three analysed tomato species (**Figure 10A, B and C**). Adding up the non-transformed LFQ-intensities for each sample showed increased total intensities in Bion-induced samples relative to mock-induced samples that mirror trends in protein concentrations assessed biochemically (compare **Figure 10D, E and F** with **Figure 6B**).

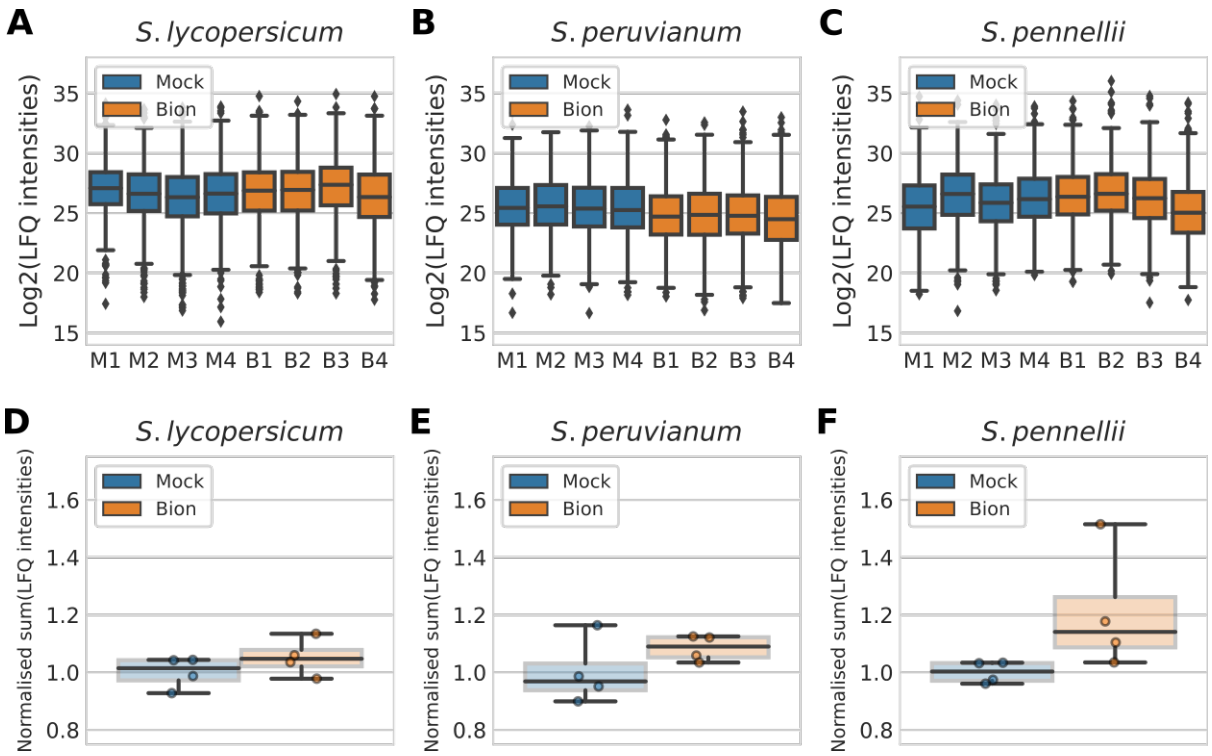


Figure 10: High-quality mass spectrometry data of tomato apoplastic fluids: (**A, B, C**) Distribution of log2-transformed LFQ-normalised values for all protein groups detected in apoplastic fluid extracts of *S.lycopersicum* (**A**), *S.peruvianum* (**B**) and *S.pennellii* (**C**). Mock- (M1-4) and Bion-treated (B1-4) samples show normally distributed log2-transformed LFQ-normalised intensities for all protein groups. (**D, E, F**) Sum of LFQ-normalised intensities for each sample of *S.lycopersicum* (**D**), *S.peruvianum* (**E**) and *S.pennellii* (**F**). Values were normalised relative to the mean of respective mock-treated controls. (**A, B, C, D, E, F**) Centre lines are the median, bounds of boxes the 25th and 75th percentiles and whiskers 1.5 x interquartile range (IQR) from 25th and 75th percentiles. Diamnods indicate outliers.

Before analysing MS data in detail, we reduced proteomics data sets to protein groups containing proteins predicted to be secreted and robustly detected in either or both mock- and Bion-induced samples. A robust detection was defined by protein groups being present in at least three out of four samples within any treatment. This filter of robust detection reduced the number of different protein groups detected in samples of *S.lycopersicum* from 3709 to 563, in *S.peruvianum* from 2805 to 499 and in *S.pennellii* from 2752 to 428. Secretion of a protein was based on prediction by either SignalP5.0, TargetP2.0 or ApoplastP1.0 (Sperschneider et al. 2018; Armenteros et al. 2019a; Armenteros et al. 2019b). This predicted secretion filter further reduced the number of different protein groups detected in samples of *S.lycopersicum* to 178, in *S.peruvianum* to 175 and in *S.pennellii* to 190. Next, we prepared the data sets to understand effects of Bion treatments within each species.

As linear dimensionality reduction via principle component analysis (PCA) requires centred data sets, we standardised intensities of protein groups around a mean of zero by subtracting the protein group's mean from all respective samples. In contrast to a z-score, which furthermore scales values, this normalisation preserves the relative differences in intensities between samples and represents induced changes between protein groups more accurately. For a comparison of centred-only and z-score standardisation, see **Figure 11**. PCA analysis of each tomato species showed that Bion treatment explains most of the

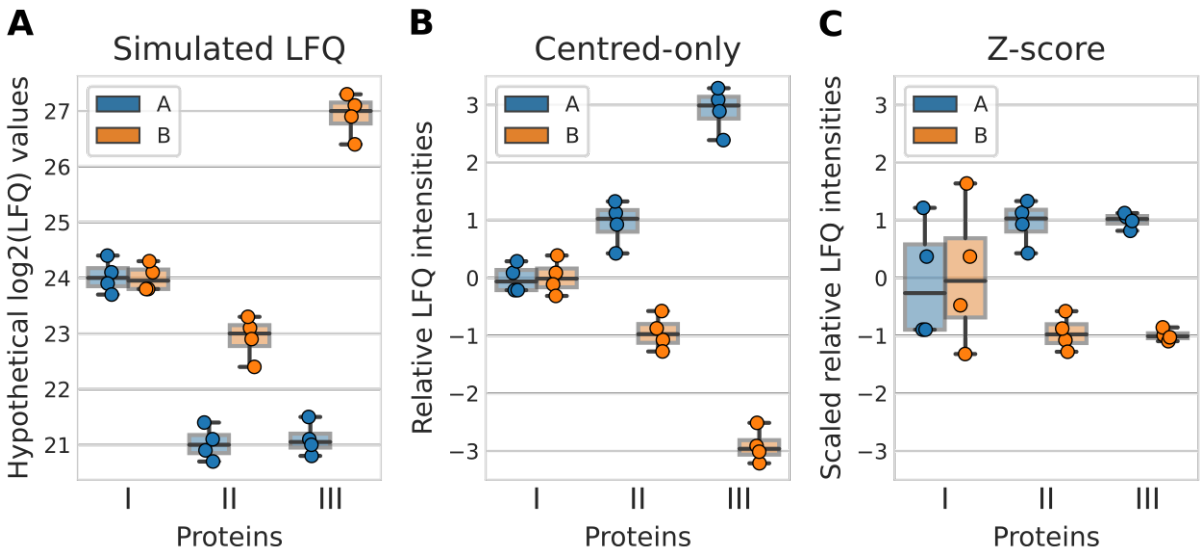


Figure 11: Comparison of standardisation methods using simulated data: (**A**) Generated, hypothetical values represent not- (I), weakly- (II) and strongly- (III) induced protein groups by treatment B. Values were standardised by subtracting each protein group's mean from values within the protein group (**B**) or by calculating the z-scores for values within a protein group (**C**).

differences observed within species (**Figure 12**). This makes us confident that observed differences in tomato AF extracts of each species are largely the result of biotic stress signalling.

2.3.3.1 Correlating PR accumulation between protein gels and MS data

To understand how Bion-induced changes in AFs compare between tomato species, we clustered proteins into orthogroups (OGs) based on whole proteomes of *S.lycopersicum,*

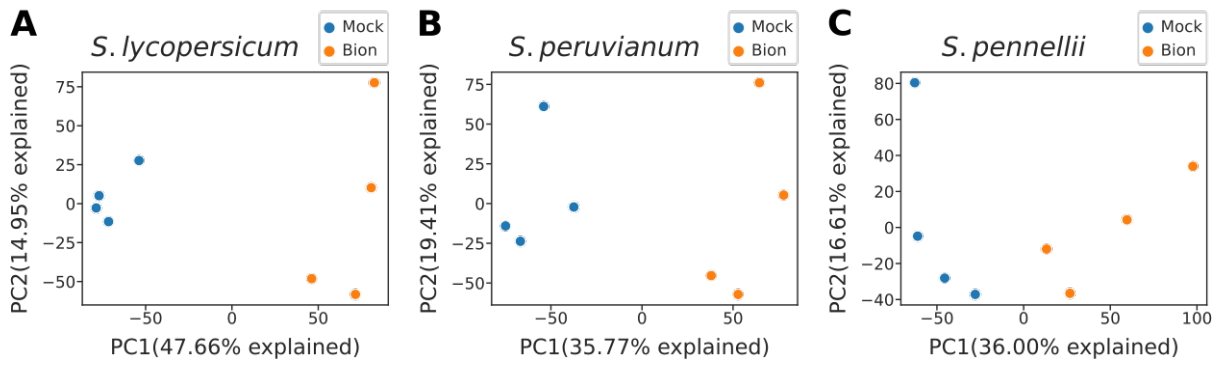


Figure 12: Bion-mediated biotic stress induction explains differences between mass spectrometry data of tomato apoplastic fluid extracts: Principal component analysis (PCA) of robustly detected, predicted to be secreted protein groups of *S.lycopersicum* (**A**), *S.peruvianum* (**B**) and *S.pennellii* (**C**) are mostly explained by Bion treatment.

S.peruvianum, S.pennellii and *N.benthamiana* using Orthofinder2.0 (Emms and Kelly 2019). We identified a total of 325 shared orthogroups with the average orthogroup containing 1.1 ± 0.6 proteins in *S.lycopersicum*, 1.2 ± 0.9 in *S.peruvianum* and 1.3 ± 1.3 in *S.pennellii*. The intensity of an orthogroup was defined by the sum of the LFQ intensities of each contributing, robustly detected and predicted to be secreted protein group. As before, we normalised log2-transformed intensities of orthogroups to be centred around zero for each species.

To validate our data set, we looked for correlation between PR proteins detected on protein gels and their relative changes in MS data. Expected biotic stress-induced PR proteins in tomato AFs are PR1 (14-17 kDa), PR2 (34-36 kDa), PR3 (22-26 kDa) and PR7 (69 kDa) (Joosten and De Wit 1989; Tornero et al. 1996; van Loon et al. 2006; Sinha et al. 2014). Our protein gels indeed displayed Bion-induced accumulation of PR proteins of similar molecular weight at bands I-IV (**Figure 13A**). To compare if LFQ intensity of orthogroups mirrored this PR protein accumulation, we selected the ten orthogroups of highest LFQ intensity of each tomato species which amounted to a total of 21 tomato orthogroups. These orthogroups encoded one Bion-induced PR7 orthogroup (band I) and two Bion-induced PR2 orthogroups (band II) (**Figure 13A**). However, PR3 (band III) and PR1 (band IV) orthogroups were not amongst the orthogroups with highest LFQ intensities (**Figure 13A**). Bion-induced 25 kDa PR3 chitinases were detected, but at

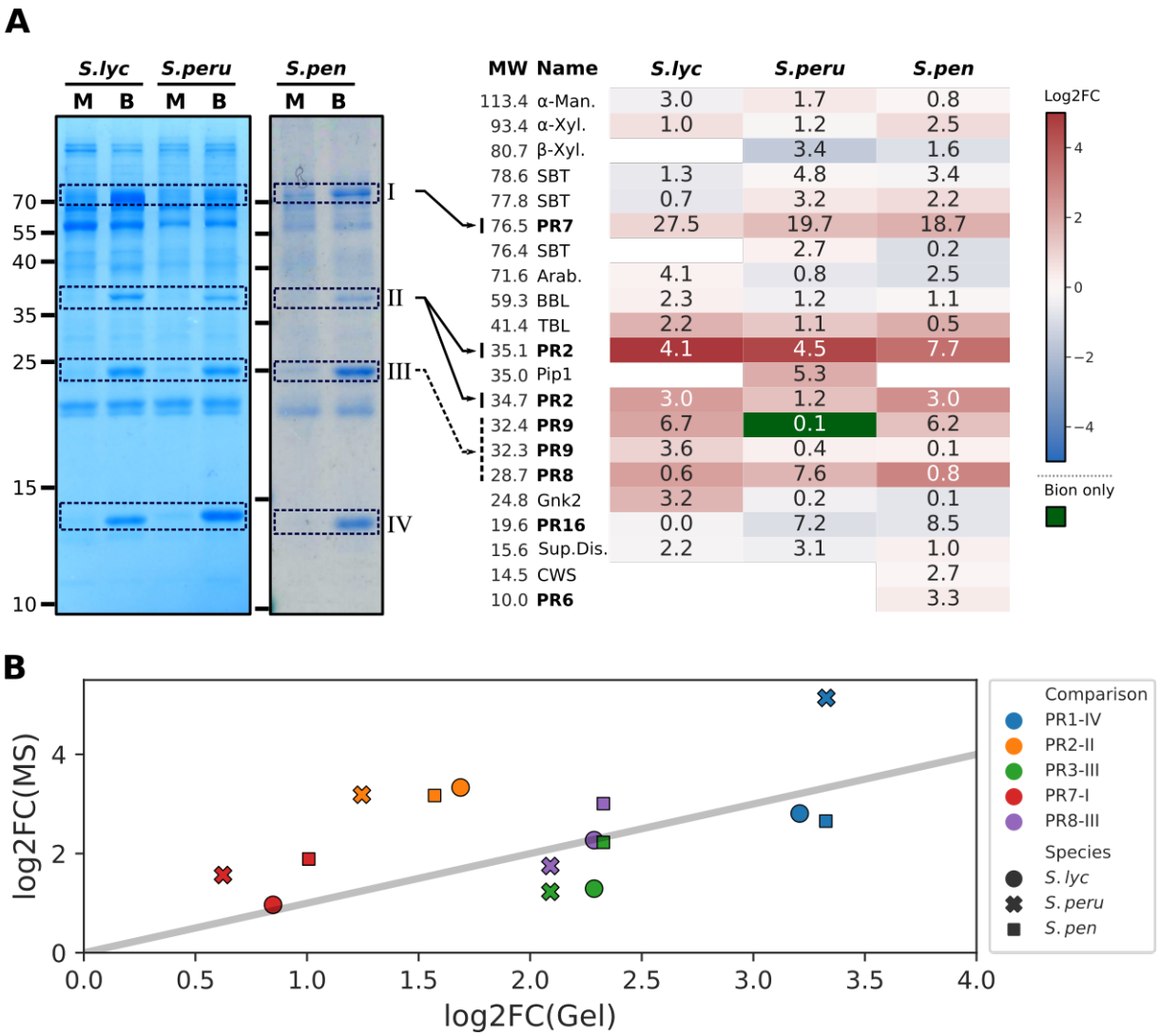


Figure 13: Orthogroup-based mass spectrometry data describes PR protein accumulation. (**A**) Combined different protein gels of apoplastic fluid (AF) from mock- (M) and Bion-treated (B) tomatoes. Dashed boxes labelled I (PR7), II (PR2), III (PR3) and IV (PR1) indicate accumulating PR proteins which were quantified using ImageJ and correlated with log2FC of MS data in (**B**) (Schneider et al. 2012). Bars on the left of the *S.pennellii* gel indicate the position of the corresponding molecular weight ruler next to the gel of *S.lycopersicum* and *S.peruvianum*. Solid arrows next to PR protein bands indicate corresponding high-intensity MS orthogroups and dashed arrows indicate Bion-induced PR proteins of comparable size but different annotation. The heatmap shows 21 Orthogroups including the ten most abundant orthogoups for each of the three tomato species. Molecular weight (MW) is in kDa and without signal peptides. Values in the heatmap represent the mean percentage share of intensity of an orthogroup in Bion-induced AFs and colours encode Bion-induced log2-fold changes (log2FC). (**B**) Correlation of Bion-induced log2FC of MS and Gel band intensities. The grey line indicates a perfect 1:1 correlation. α-Man.: α-mannosidase, α/β-Xyl: α/β-xylosidase, SBT: subtilase-like, Arab.: arabinofuranosidase, BBL: berberine-bridge-like, Sup.Dis: superoxide dismutase, CWS: cell-wall structural protein, *S.lyc*: *S.lycopersicum*, *S.peru*: *S.peruvianum*, *S.pen*: *S.pennellii*.

low shares of LFQ intensities across all three species (0.1-0.7%). Band III represents a PR protein with chitinase activity in tomato (Joosten and De Wit 1989), which could correspond to PR8 class III chitinases which has a comparable molecular weight and is Bion-induced. The inconsistent high intensity of the PR8 orthogroup across tomato species, however, makes it unlikely to represent the uniformly accumulating protein band of the three tomato species (**Figure 13A**). Alternatively, PR9 peroxidases are Bion-induced and have a similar molecular weight. For band IV, we did not detect a Bion-induced orthogroup with high LFQ intensity (**Figure 13A**), but we detected Bion-induced PR1 orthogroups across the three tomato species at low shares of LFQ intensities (0.1-0.8%). We conclude that the low share of LFQ intensities of PR1 and PR3 did not reflect their high abundance on protein gels in the literature.

Biases in quantification of LFQ intensities could stem from different ionisation efficiencies, glycosylation of peptides or lack of generation of distinct tryptic peptides (Müller and Winter 2017; Dincer et al. 2022). To investigate if insufficient different tryptic peptides were detected for PR1s and PR3, we predicted all tryptic peptides and determined their identification for each species. We used full length protein sequences for tryptic peptide prediction with unipept v3.0.2 (Verschaffelt et al. 2021), and the number of predicted tryptic peptides (>6 amino acids) of PR proteins is described by the denominators in **Table 7**. A tryptic peptide was considered identified if the full peptide, a semi-tryptic peptide subset or a peptide with a missed tryptic cleavage site containing the full predicted tryptic peptide were identified in MS. The identification workflow was standardised in the script 'peptides_overview.sh' which is deposited in our GitHub repository (Homma 2022d). For PR1s, we identified between 12.5% (1/8) and 62.5% (5/8) of all predicted tryptic peptides and for each tomato species, at least 37.5% of predicted tryptic peptides were detected for one PR1 orthogroup (**Table 3**).

For 25 kDa PR3, we identified between 57.1% (6/11) and 72.7% (8/11) of all predicted tryptic peptides in MS data (**Table 3**). To put these fractions of identified tryptic peptides into context, for 35.1 kDa PR2, we identified between 26.6% (4/15) and 40.0% (6/15)

Table 3: Fraction of predicted tryptic peptides identified in MS data

Orthogroup	*S.lycopersicum*	*S.peruvianum*	*S.pennellii*
PR1 (OG0003255)	5/8	-	4/7
PR1 (OG0017254)	3/8	-	3/8
PR1 (OG0017254)	-	1/8	-
PR1 (OG0020546)	3/7	4/7, 4/7	-
PR3 (OG0022181)	6/11	8/12	8/11
PR2 (OG0003427)	5/15	4/15	6/15
PR7 (OG0000167)	18/35	16/36	20/36

of predicted tryptic peptides and for PR7, we identified between 44.4% (16/36) and 55.5% (20/36) of predicted tryptic peptides (**Table 3**). We thus detected a comparable number of different tryptic peptides for both PR1 and PR3 compared to high-intensity PR orthogroups. We concluded that low LFQ intensity of PR1s and PR3 was not due to insufficient different tryptic peptides identified.

To understand correlation between Bion induction of PR proteins in MS data compared to protein gels, we quantified protein accumulation from protein gels using ImageJ and correlated increasing signal intensities of protein bands with increasing LFQ intensities of PR1, PR2, PR3, PR7 and PR8. To calculate PR protein induction in MS data, we added up non-transformed LFQ intensities of protein groups detected in orthogroups annotated as respective PR proteins and then calculated their log2-fold changes (log2FC). First, there is a good correlation between band I and PR7 and we consistently detected higher induction in MS- compared to gel intensity for band II and PR2 (**Figure 13B**). We also found a good correlation between Bion induction in MS- compared to gel intensity between band III and the chitinases PR3 and PR8 (**Figure 13B**). For band IV and PR1, Bion-induction in MS- compared to gel intensity was comparable in *S.lycopersicum* and *S.pennellii* and higher in *S.peruvianum* and (**Figure 13D**). Taken together, MS data reflects information of differential PR protein accumulation and shares of LFQ intensity reflect high abundance of some, but not all PR proteins. This made us confident that an orthogroup-based approach is robust enough to identify Bion-induced changes in AF from different species.

2.3.3.2 Bion-induced changes in tomato secretomes

To explain differences and similarities between species and treatments, we performed a PCA analysis and calculated pairwise pearson correlations between MS samples of centred orthogroup intensities (**Figure 14**). In PCA analysis, replicates from the same species and treatment cluster close together (**Figure 14B**). The two components explaining most

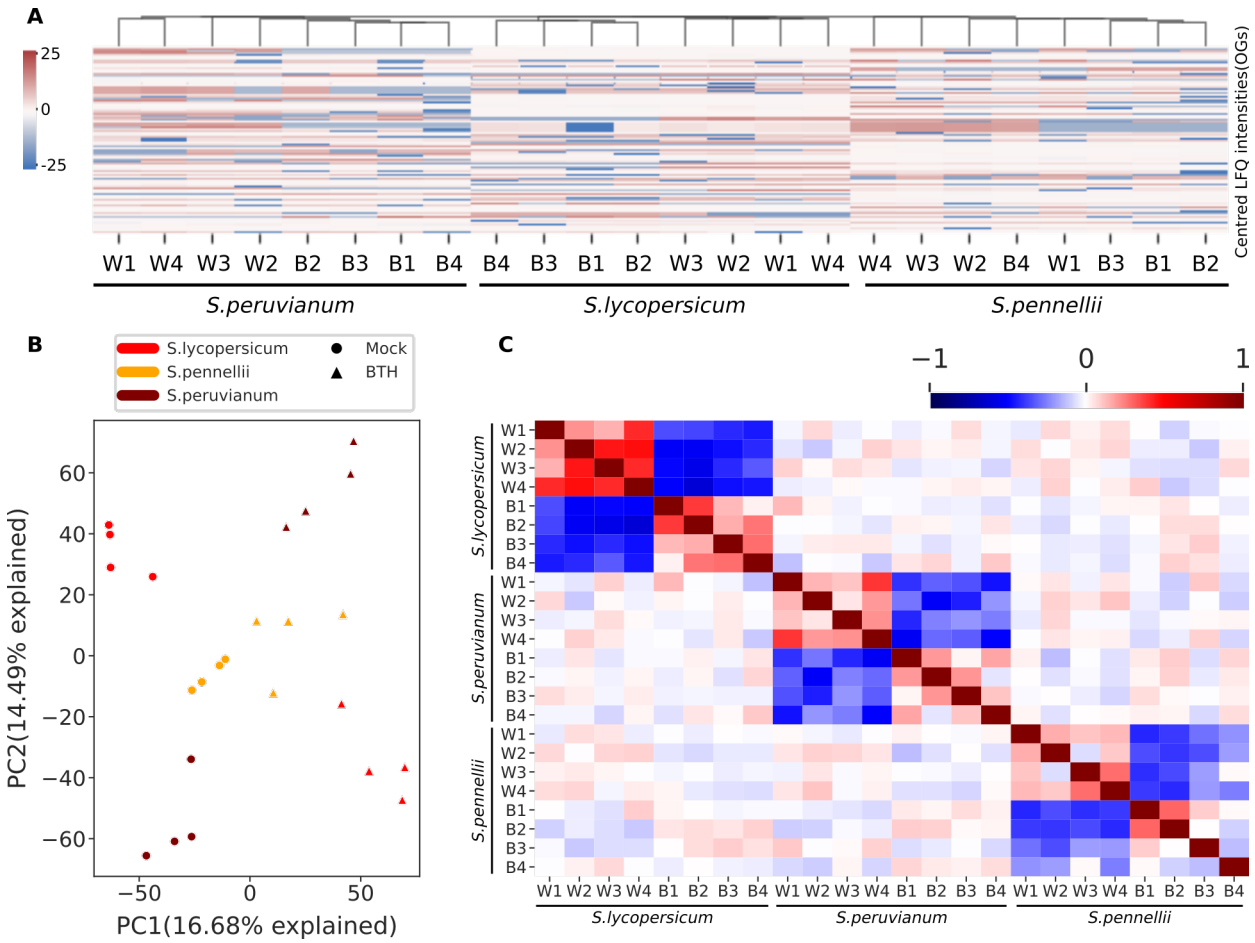


Figure 14: Orthogroup-based analysis of proteomic data of tomato AFs shows similarities within and dissimilarities between tomato species. (**A**) Clustermap of centred LFQ intensities of orthogroup-based tomato MS data. Clusters were based on Jaccard distances between mock-/water- (W) and Bion-treated (B) samples. Data visualised in (**A**) was used to for analysis in (**B**) and (**C**). (**B**) Principle component analysis (PCA) of centred, orthogroup-based tomato AFs shows that each tomato species clusters based on different treatments. (**C**) Pairwise pearson correlation matrix shows similarities and differences between centred mock-/water- (W) and Bion-treated (B) samples of different tomato species.

of the difference between all 24 samples were explained by differences between treatments (**Figure 14B** PC1) and by differences between species (**Figures 14B** PC2). Here, treatment explains less of the differences between samples compared to previous PCAs within

species (**Figure 12**), which indicates that the three tomato species respond differently to Bion treatment.

To understand similarities between all samples, we measured linear correlation between all tomato AF samples by calculating the pearson correlation coefficient for each pairwise comparison. Within each tomato species, there is positive correlation between most samples of the same treatments (**Figure 14C**). This indicates that many orthogroups contain similar centred values between samples of the same species and treatment. By contrast, most samples between treatments and within species showed negative correlation (**Figure 14C**). This indicates that many orthogroups contain contrasting values between differently treated samples of the same species. Low correlation between samples within the same treatment and species could stem from few orthogroups with missing values in individual samples. Interestingly, samples from different species demonstrated little correlation to each other, which showed that regardless of treatment, centred values of orthogroups do not correlate strongly between different tomato species (**Figure 14C**).

2.3.3.3 LFQ intensities explain similarities between tomato immuno secretomes

To understand similarities between apoplasts of biotic stress-induced tomatoes, we analysed each orthogroup's relative share of total untransformed, LFQ intensity for each sample to derive a proxy for their protein concentration. This assumption is based on the conclusion that MaxLFQ-derived LFQ values, a label-free quantification score implemented in the MaxQuant MS data analysis software, is a proxy for absolute protein abundance, despite potential quantification biases (Cox and Mann 2008; Cox et al. 2014). To estimate Bion-induced changes, we compared all induced orthogroups. In *S.lycopersicum*, 24 orthogroups were induced by Bion treatment, in *S.peruvianum* 17 and in *S.pennellii* 14. Of these, only five orthogroups were induced in all three species (**Figure 15A**). These five orthogroups were highly-abundant in Bion-induced AFs, as they defined 41.8%, 33.0% and 36.3% share of LFQ intensities in *S.lycopersicum, S.peruvianum* and *S.pennellii* respectively (**Figure 15B**). The remaining induced orthogroups combined for only 13.1% in

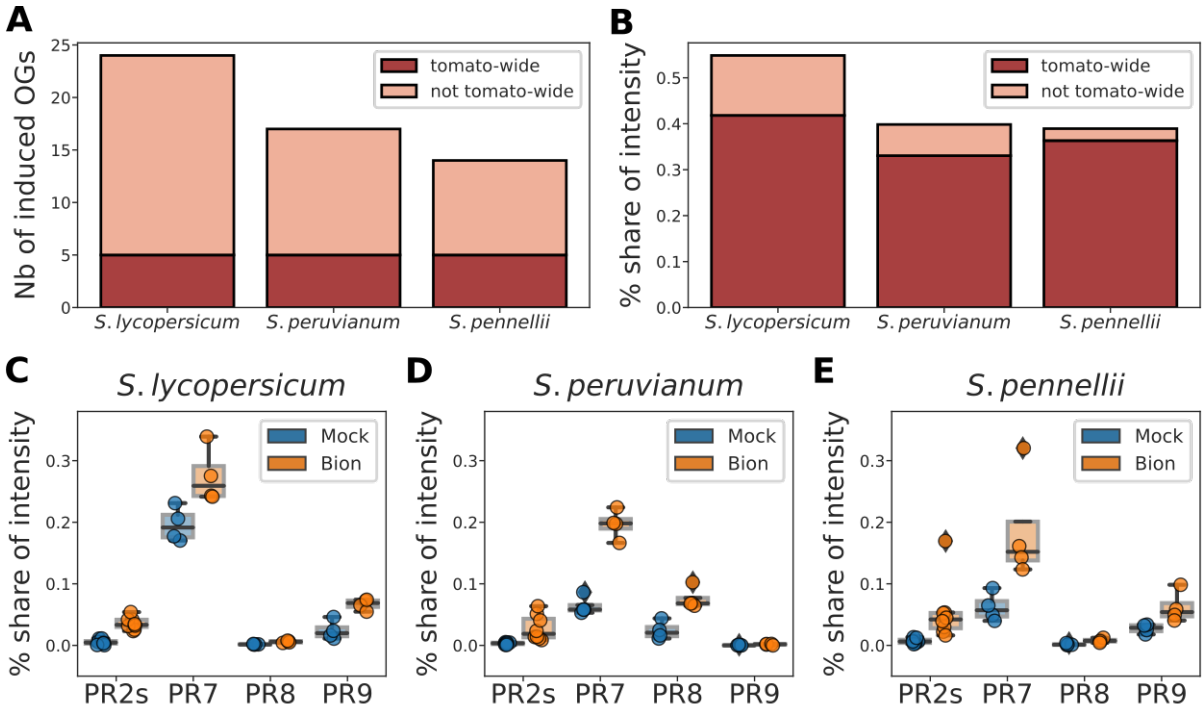


Figure 15: Few, shared Bion-induced orthogroups dominate tomato apoplasts. (**A, B**) Orthogroups induced by Bion in all three tomato species are labelled as 'tomato-wide'. (**A**) Total number of Bion-induced orthogroups per species. (**B**) Relative share of total LFQ intensities of Bion-induced orthogroups per species. (**C, D, E**) Total percentage share of LFQ intensities of tomato-wide induced PR proteins in *S.lycopersicum* **C**, *S.peruvianum* **D** and *S.pennellii* **E**. Centre lines are the median, bounds of boxes the 25th and 75th percentiles and whiskers 1.5 x interquartile range (IQR) from 25th and 75th percentiles. Diamonds represent outliers.

S.lycopersicum, 6.8% in *S.peruvianum* and 2.6% in *S.pennellii*. The five highly-abundant shared orthogroups encoded two PR2s, PR7, PR8 and PR9. Of these five PR proteins, PR7 (P69 subtilases) is the most abundant orthogroup in all three species (**Figure 15C, D, E**). PR7 was significantly more abundant in Bion-induced AFs of *S.lycopersicum* (27.5%) than in *S.peruvianum* (19.7%), but not significantly more than in *S.pennellii* (18.7%), likely due to higher variation within *S.pennellii* (respective FDR-corrected p-values of one-sided, unpaired t-tests are 0.023 and 0.067). In contrast, induction of PR7 was higher in wild tomato at 2.9-fold increase in *S.peruvianum* and 3.7-fold increase in *S.pennellii* compared to 1.96-fold increase in *S.lycopersicum*.

Technically, all Bion-induced orthogroups could be classified as containing PR proteins (van Loon et al. 2006). The five shared orthogroups, however, contained highly-abundant proteins in tomato AFs that are conserved in their responsiveness to biotic stress across

distantly related tomato species. As a consequence, these highly abundant proteins will be encountered frequently and are to be overcome by plant pathogens attempting to colonise the tomato apoplast. These five orthogroups thus highlight strong similarities between different tomato apoplastic proteomes. In contrast, the remaining induced orthogroups, highlight non-conserved responses to biotic stress. While large in the number of orthogroups, these orthogroups encode proteins without evidence for high abundance.

2.3.4 Positive selection in biotic stress-induced tomato secretomes

To add an additional line of evidence for proteins to be implicated in plant immunity, we analysed the molecular evolution in tomato genes encoding secreted, Bion-induced proteins. We selected twelve tomato species for which whole genomes have been sequenced (100 Tomato Genome Sequencing Consortium 2014; Lin et al. 2014; Stam et al. 2019b; Wang et al. 2020b). These twelve species represent the genomic breadth and diversity of the four tomato species clades *Lycopersicon*, *Arcanum*, *Eriopersicon* and *Neolycopersicon*.

2.3.4.1 Inferring ploidy levels of twelve tomato species

To know how many different potential alleles to expect per gene, we inferred the ploidy levels of each tomato species. We inferred their ploidy levels by analysing coverages of heterozygous kmers using smudgeplot (**Figures 16, 17, 18**) (Ranallo-Benavidez et al. 2020). Smudgeplot identifies kmers that differ by exactly one nucleotide and compares the number of reads identified for each sequence to infer the most likely ploidy of the genome. We confirmed diploid genomes for all twelve tomato species.

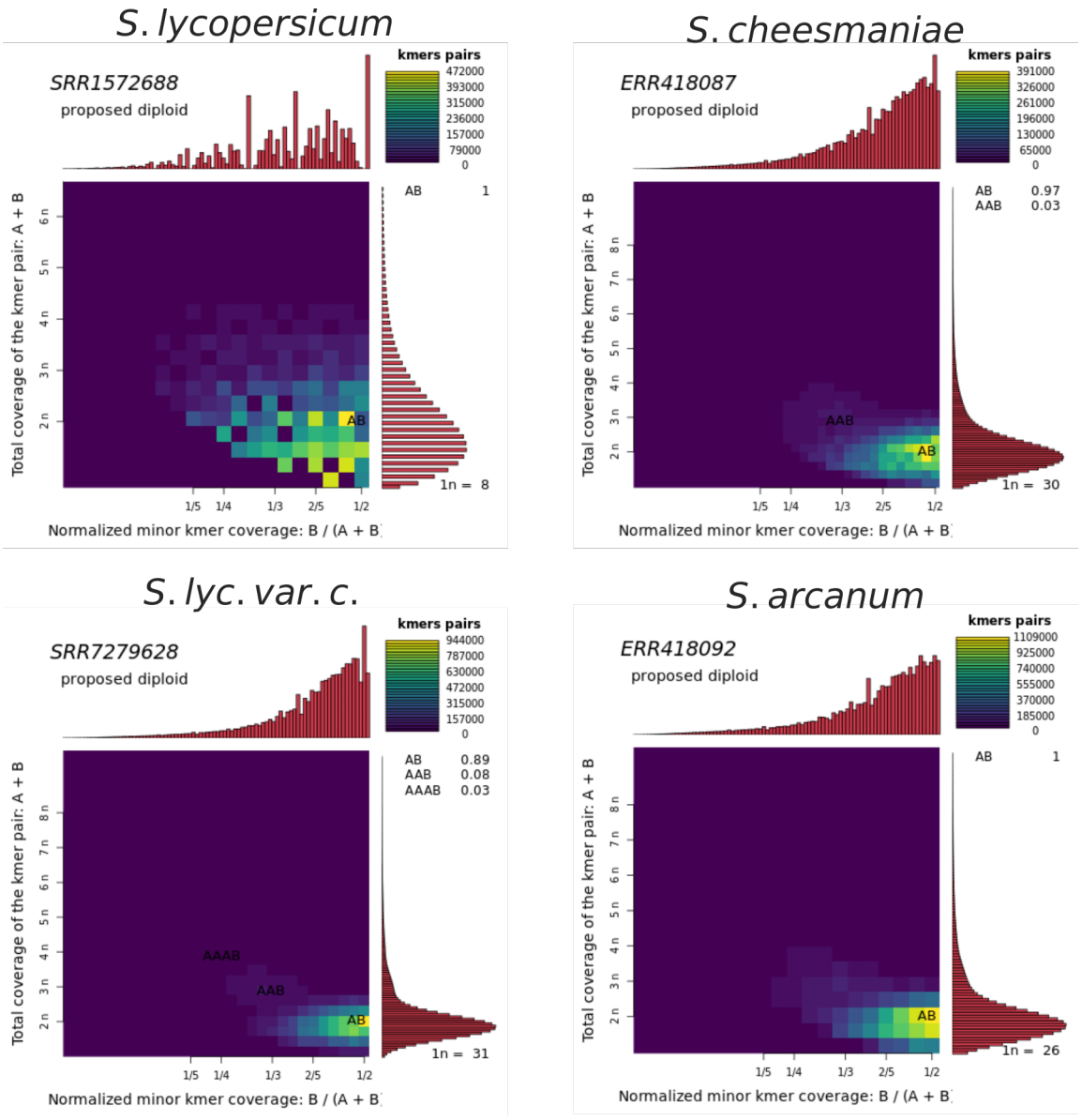


Figure 16 All tomato species have diploid genomes. Inferred ploidies based on analysing heterozygous kmers and comparing each combined coverage (y-axes) to their respective relative coverage (x-axes) using smudgeplot (Ranallo-Benavidez et al. 2020). SRR and ERR references refer to NCBI's sequence read archive sequencing runs of whole-genome sequencing data used for all analyses in this section of the chapter. *S. lyc.* var. *c.* stands for cherry tomato *Solanum lycopersicum* var. *cerasiforme*.

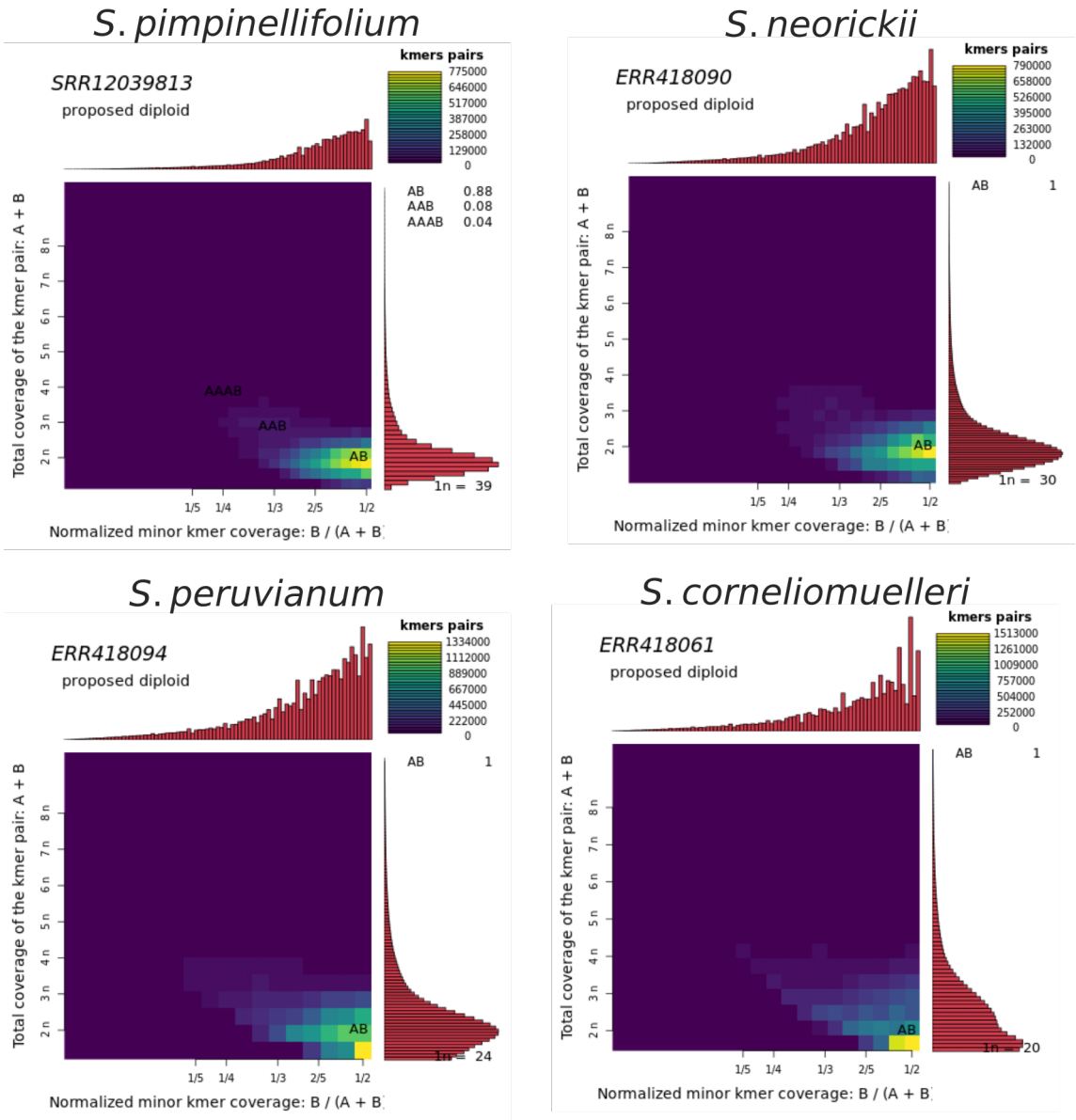


Figure 17 All Tomato species have diploid genomes. Inferred ploidies based on analysing heterozygous kmers and comparing each combined coverage (y-axes) to their respective relative coverage (x-axes) using smudgeplot (Ranallo-Benavidez et al. 2020). SRR and ERR references refer to NCBI's sequence read archive sequencing runs of whole-genome sequencing data used for all analyses in this section of the chapter.

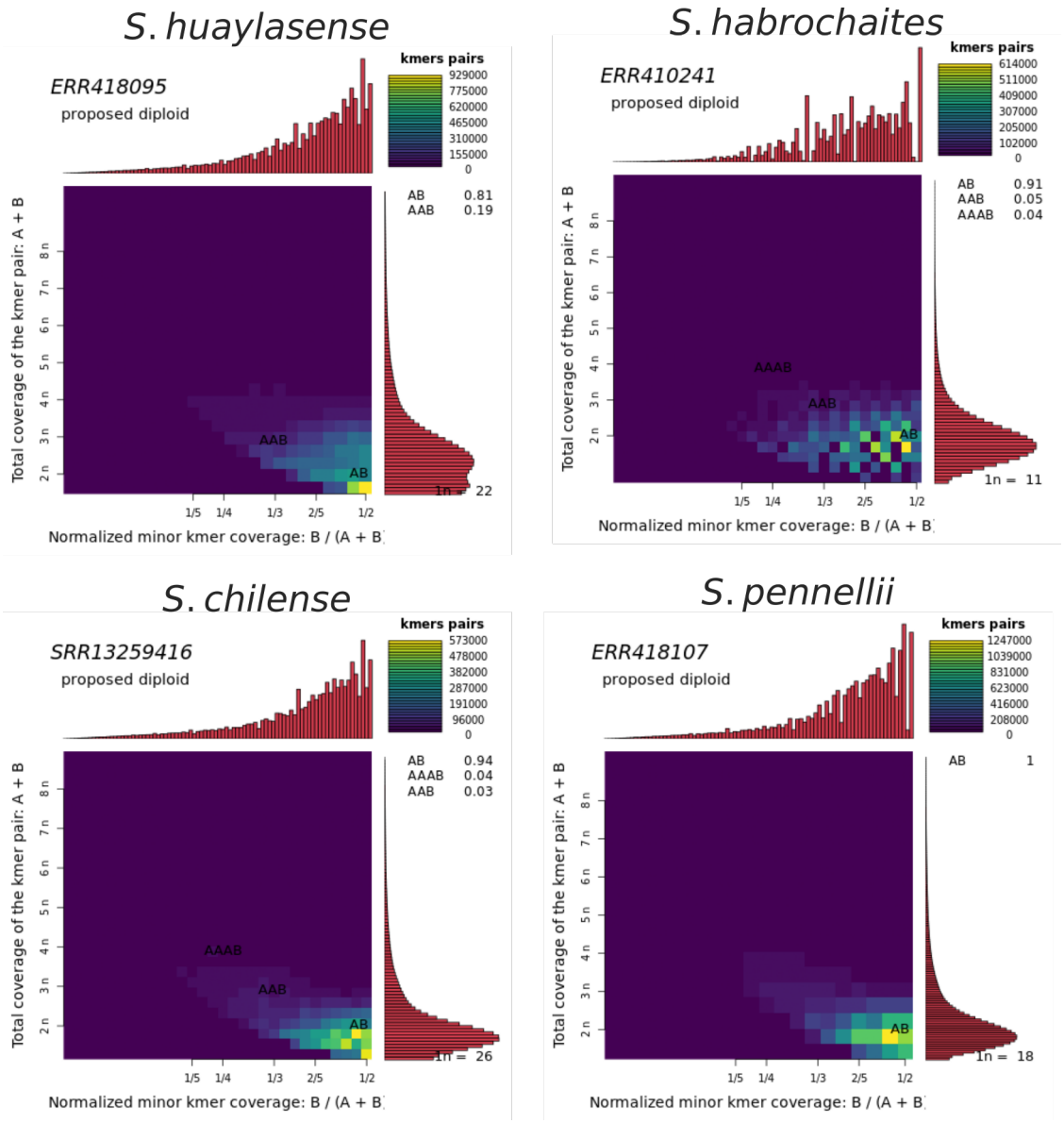


Figure 18 All Tomato species have diploid genomes. Inferred ploidies based on analysing heterozygous kmers and comparing each combined coverage (y-axes) to their respective relative coverage (x-axes) using smudgeplot (Ranallo-Benavidez et al. 2020). SRR and ERR references refer to NCBI's sequence read archive sequencing runs of whole-genome sequencing data used for all analyses in this section of the chapter.

2.3.4.2 Detecting positive selection in Bion-induced tomato proteins

In the absence of genome assemblies of all twelve tomato species, we mapped genomic reads from eleven tomato species against the Sol4.0 *S.lycopersicum* reference genome and used information of sequencing reads spanning multiple SNPs and small InDels to phase variants. We then selected loci from *S.lycopersicum* encoding proteins from the 39 biotic stress-induced orthogroups. 38 of the 39 orthogroups contained at least one homolog in *S.lycopersicum* and these covered 51 genomic loci. The loci of these *S.lycopersicum* homologs were used to generate alleles from wild tomato species using their genomic reads mapped against the *S.lycopersicum* reference genome. For all phased unique alleles, we constructed phylogenetic trees to analyse each locus for site-wise positive selection. The workflow is conceptually described in **Figure 19** and consists of pipelines for variant calling, phasing, allele generation and positive selection, made by combining several bioinformatics tools. The pipelines can be found in our GitHub repository (Homma 2022g; Homma 2022e; Homma 2022c; Homma 2022f). We detected positive selection in 21 of 51 of the genes encoding apoplastic Bion-induced tomato proteins (41.2%). This includes three proteins that were previously reported to contain variant residues or to be under positive selection: class I chitinase (PR3) is under positive selection in *Arabis* and *Solanaceae* species (Bishop et al. 2000), Pip1 contains high levels of amino acid diversity of surface-exposed residues in tomato and potato alleles (Shabab et al. 2008; Hörger and van der Hoorn 2013) and P69B (PR7) has high amino acid variation of residues surrounding the substrate binding groove between tomato species (Hörger and van der Hoorn 2013). To put our results into context, comparable positive selection analyses were conducted for 19,135 orthologous proteins from seven different strawberry species which identified that 3.8-11.1% of these proteins are under positive selection (Qiao et al. 2021). To compare the fraction of genes under positive selection, we corrected p-values of our positive selection not for 51, but 19,135 analyses as in Qiao et al. 2021. This resulted in 5 of 51 genes (9.8%) identified under positive selection which is within the range of genome-wide positive selection in strawberry. These five genes showed traces of positive selection strong enough to be identified in a genome-wide comparison. In descending order of statisti-

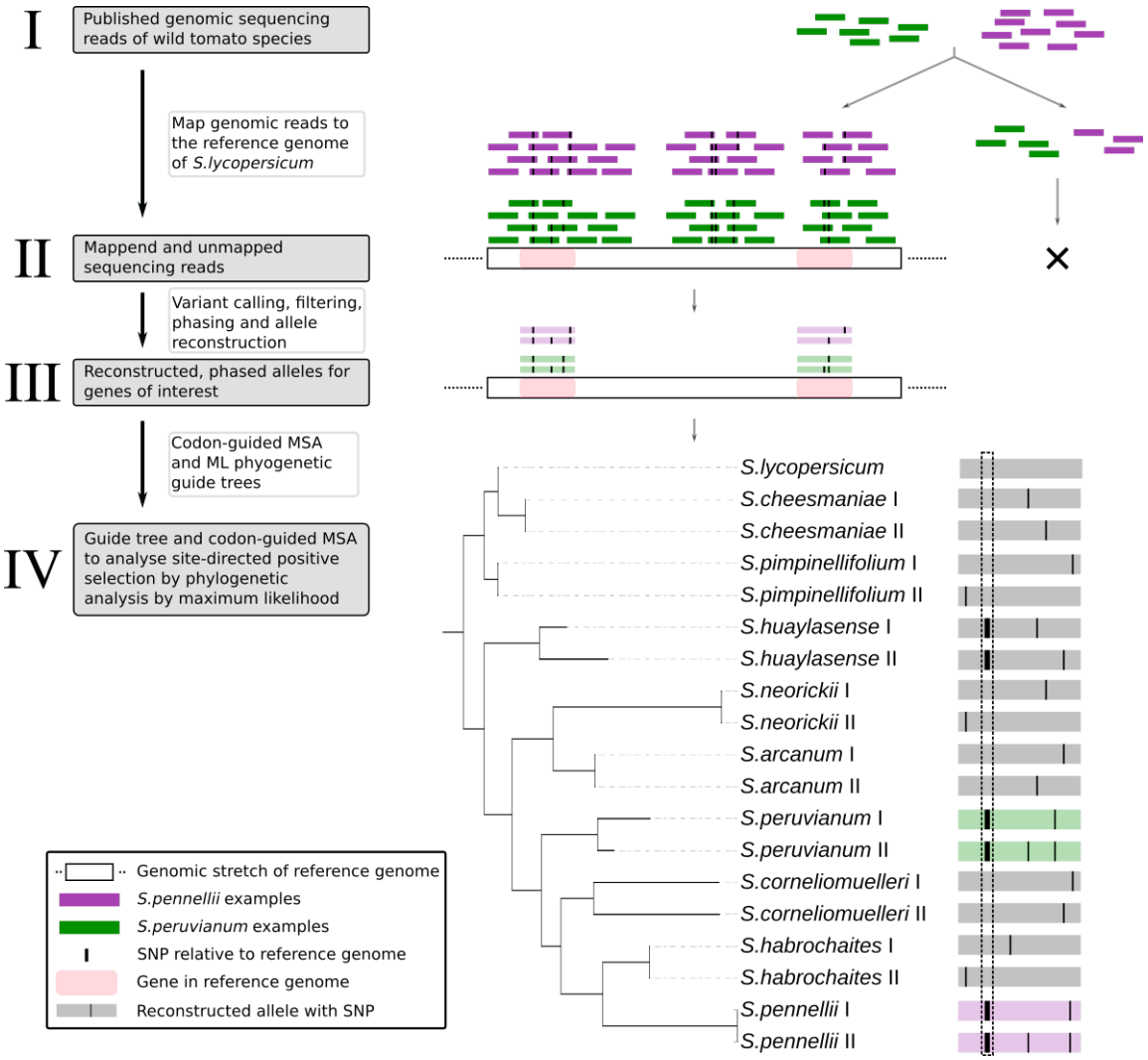


Figure 19 Example workflow to analyse site-directed positive selection in wild tomato species. Published, quality-trimmed genomic sequencing reads of wild tomatoes (**I**) were mapped against the Sol4.0 *S.lycopersicum* reference genome (**II**), variants were called, phased and differing alleles from loci of interest generated (**III**). Maximum likelihood phylogenetic trees were generated using different tomato alleles from loci of interest and site-directed positive selection analysed (**IV**). The dashed box in (**IV**) highlights an example site that independently accumulated a mutation several times.

cal significance, these five genes encode Solyc02g077040.4.1 (Pip1), Solyc03g044160.3.1 (Malectin-like protein kinase), Solyc03g082920.4.1 (HSP70), Solyc08g079870.3.1 (P69B) and Solyc08g007670.1.1 (SBT1.10q) (Reichardt et al. 2018). The first gene not exceeding the significance threshold after proteome-wide multiple testing correction was Solyc01g105070.3.1 (Peroxidase) with an uncorrected p-value of 0.000007. Reducing the number of positive selection analyses by selecting Bion-induced proteins thus enabled identifying genes under weaker positive selection compared to genome-wide analyses. An

overview of proteins encoded by genes under positive selection from Bion-induced orthogroups can be found in **Table 4**. Evidence of positive selection acting on Pip1, P69B

Table 4: Bion-induced apoplastic proteins under positive selection

Species-specific MS data								Tomato-wide Pos. Selection		
Orthogroup	**Funct. Annot.**	**Log2FC**			**% Intensity**			**Slyc ID**	ω	ρ
Proteases		**A**	**B**	**C**	**A**	**B**	**C**			
OG0019347	S10 carboxypep.	6.0			0.02			**Solyc06g083020.3.1**	2.1	0.08
OG0011969	Pip1		3.0			5.27		**Solyc02g077040.4.1**	8.5	0.06
OG0000167	S8 Subtilases	1.3	2.6	2.5	27.48	19.63	18.69	**Solyc08g007670.1.1**	29.9	0.00*
OG0000167	S8 Subtilases	1.3	2.6	2.5	27.48	19.63	18.69	Solyc08g079850.3.1	2.1	0.17
OG0000167	S8 Subtilases	1.3	2.6	2.5	27.48	19.63	18.69	Solyc08g079860.2.1	6.0	0.01
OG0000167	S8 Subtilases	1.3	2.6	2.5	27.48	19.63	18.69	**Solyc08g079870.3.1**	2.1	0.18
OG0000167	S8 Subtilases	1.3	2.6	2.5	27.48	19.63	18.69	Solyc08g079880.2.1	2.4	0.20
Protease Inhibitors										
OG0009032	Cys-protease Inh.		6.0	6.0		0.02	0.02	Solyc05g054120.1.1	24.6	0.02
OG0005663	Protease Inh.	2.0			0.05			**Solyc03g020030.3.1**	3.8	0.09
Glycosylhydrolases										
OG0015230	Chitinase Class I			6.0	0.17		0.01	**Solyc09g098540.3.1**	3.4	0.09
OG0003011	Cellulase	1.7			0.27			**Solyc08g066810.3.1**	2.9	0.11
OG0013080	GH3	1.3	1.1		0.72	0.19	0.49	Solyc04g072860.3.1	2.5	0.14
OG0022181	Chitinase Class I	2.9	2.2	2.6	0.73	0.12	0.26	Solyc02g082920.4.1	3.5	0.07
ROS-related										
OG0005350	Berberine Oxidase	1.5			0.86	0.57	0.55	**Solyc02g022850.1.1**	2.1	0.10
OG0012708	Haem Peroxidase	1.4	1.3		3.56	0.36	0.11	**Solyc01g105070.3.1**	8.1	0.01
Other										
OG0012402	MB Kinase[1]	6.0			0.06			Solyc03g044160.3.1	4.3	0.09
OG0011522	Pro-rich Protein		6.0	1.3	0.34	0.08	0.30	**Solyc12g009650.2.1**	5.4	0.01
OG0019831	FAP1[2]	3.0			0.42			**Solyc07g045440.1.1**	2.4	0.03
OG0000480	HSP70		2.0			0.17	0.01	Solyc03g082920.4.1	7.4	0.03
OG0001159	Non-specific LTP[3]			1.9			0.22	Solyc03g034338.1.1	7.2	0.13
OG0008189	Extensin-like LRR	1.1	1.8	1.9	1.91	0.50	0.78	**Solyc02g064940.1.1**	2.6	0.08

Common names for S8 subtilases are SBT1.10q (Solyc08g007670.1.1), P69B (Solyc08g079870.3.1), P69C (Solyc08g079860.2.1), P69D (Solyc08g079850.3.1) and P69J (Solyc08g079880.2.1) (Reichardt et al. 2018). Gene strings of proteins highlighted in bold were selected for cloning. A: *S.lycopersicum*. B: *S.peruvianum*. C: *S.pennellii*. ρ: Fraction of codons with ω. * Values were below the displayed threshold. [1] MB Kinase: Malectin-like binding, kinase domain. [2] FAP1: Fasciclin-like arabinogalactan protein. [3] Non-specific LTP: non-specific lipid transfer protein.

and a peroxidase (Solyc01g105070.3.1) will be explained in the next sections.

2.3.4.3 Solyc02g077040.4.1 - Pip1

The mature sequence of the immunity-related PLCP Pip1 was modelled with the swiss model online tool (Waterhouse et al. 2018a) using the *Ricinus communis* cysteine endopeptidase 1s4v.1A as a template (Than et al. 2004). Pip1 has 15 codons under positive selection (**Figure 20B**). Only residue V was in proximity to the substrate binding groove in the 3D model (**Figure 20B, residue V**). Position V has either aspartic acid, serine or glycine residues which have substantially different biochemical properties. Aspartic acid is sterically the most challenging of the three amino acids and, at apoplastic pH, will

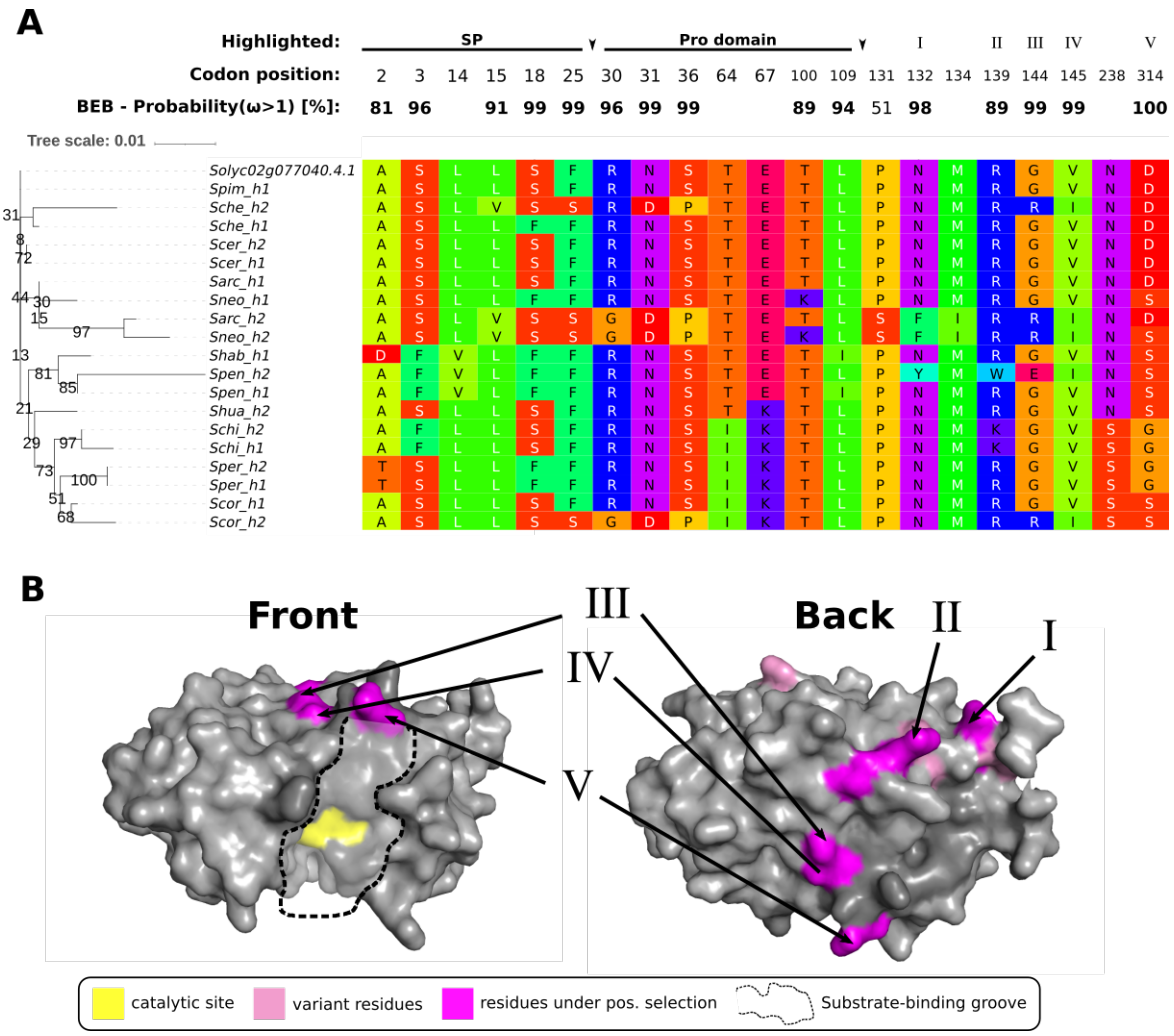


Figure 20 Combined proteomics and molecular evolution data identifies the immunity-related Papain-like cysteine protease Solyc02g077040.4.1 (Pip1) as a positive control. (**A**) Summary of positive selection analysis comprising the nucleotide-based maximum likelihood phylogenetic tree with bootstrap values and multi sequence alignments of all variant amino acid residues. SP indicates the SignalP predicted signal peptide and Bayes empirical Bayes (BEB) indicates the posterior probability of codon positions to be under positive selection. (**B**) 3D model of Solyc02g077040.4.1 based on the *Ricinus communis* cysteine endopeptidase template 1s4v.1.A.

be predominantly negatively charged. Serine exposes a small, polar and uncharged side chain and glycine has no side chain. These three different residues provide a range of biochemical environments near the substrate binding groove that could affect protein-protein interactions. The remaining four surface-exposed residues under positive selection stretch over the back of the enzyme and do not resemble the classic 'ring of fire' assumed to be caused by antagonistic interactions with cognate inhibitors (Bishop et al. 2000; Bishop et al. 2005). More surface-exposed variant residues not under positive selection were located

on the back of Pip1's substrate binding groove (**Figure 20B**).

2.3.4.4 Solyc08g079870.3.1 - P69B

The PR7 subtilase P69B was modelled using tomato SBT3 subtilase 3i6s.1.A as a template (Ottmann et al. 2009; Waterhouse et al. 2018b). Three residues under selection were close to the catalytic site (**Figure 21A**). Especially position III was highly diverse having

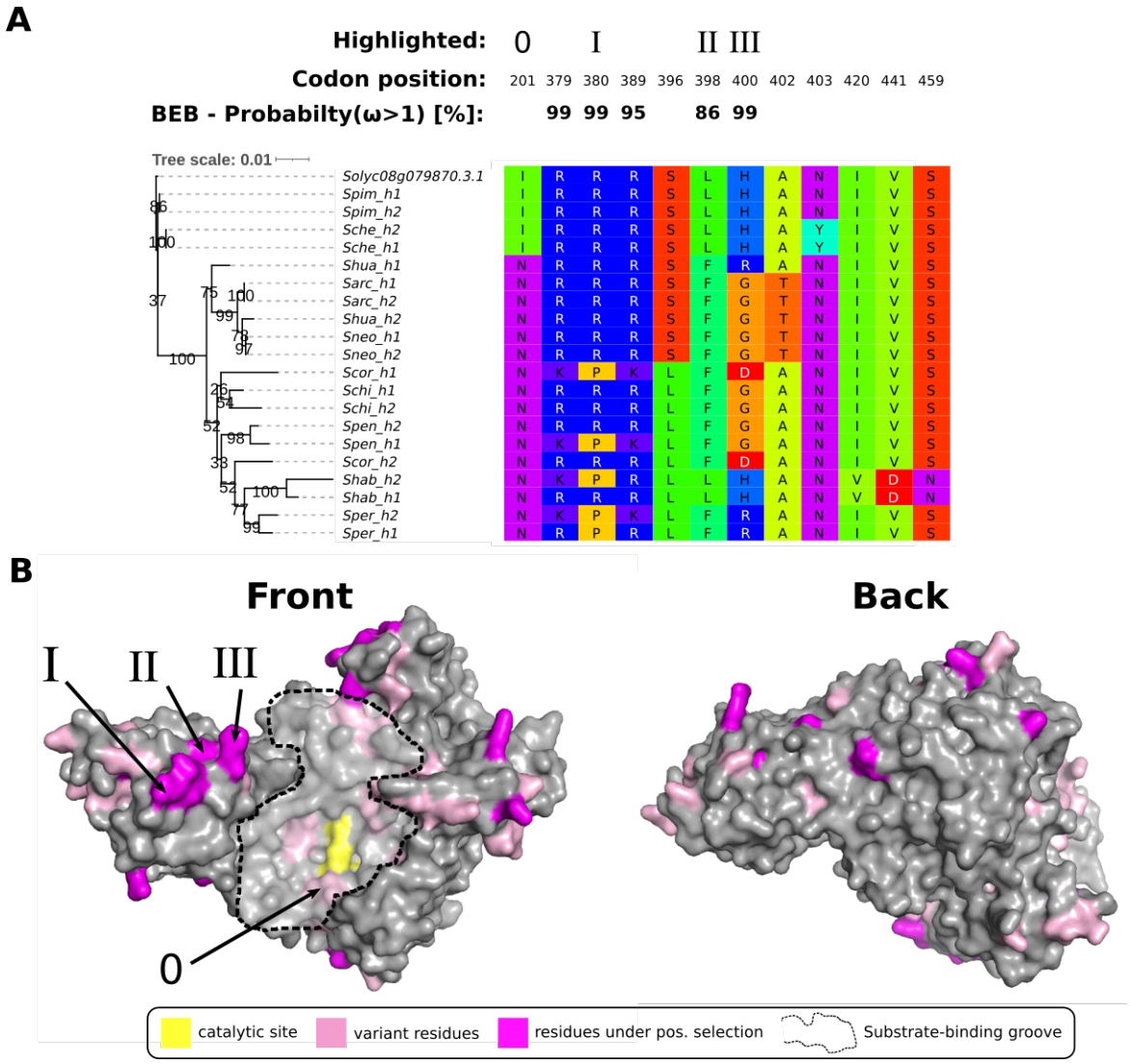


Figure 21 Combined proteomics and molecular evolution data identifies the immunity-related Papain-like cysteine protease Solyc08g079870.3.1 (P69B) as a positive control. (**A**) Summary of positive selection analysis comprising the nucleotide-based maximum likelihood phylogenetic tree with bootstrap values and multi sequence alignments of a subset of variant amino acid residues. Bayes empirical Bayes (BEB) indicates the posterior probability of codon positions to be under positive selection. (**B**) 3D model of Solyc08g079870.3.1 based on the tomato SBT3 subtilase template 3i6s.1.A. A subset of residues identified under positive selection are highlighted in magenta.

large positively charged (His and Arg), negatively charged (Asp) or small (Gly) residues (**Figure 21 A III**). Furthermore, codons at positions II and III appear to co-evolve as whenever position III is a histidine, position II is a leucine. When position III is an arginine, glycine or aspartic acid, site II is a phenylalanine. Both phenylalanine and histidine contain an aromatic side chain with a benzene ring or imidazole, respectively. Either combination thus provides an aromatic side chain at the edge of the substrate binding groove. Position I is either a large, polar residue (arginine) or cyclical and nonpolar (proline). Either has different biochemical properties and could affect protein-protein interactions. The three sites under positive selection next to the substrate binding groove are reminiscent of the 'ring of fire', observed in antagonistic enzyme-inhibitor interactions (Bishop et al. 2000; Bishop et al. 2005). Many additional variant residues, such as I201N, are located in the substrate binding groove and create different biochemical properties (**Figure 21**). While these different environments could alternate P69B inhibition, phylogenetic analyses suggest a clade-specific variation and not repeated accumulation of mutations, indicative of antagonistic arms races.

2.3.4.5 Solyc01g105070.3.1 - class III peroxidase

Solyc01g105070.3.1 is functionally annotated to be a class III haem peroxidase (PFAM PF00141 and InterPro IPR002016). Its increase in abundance during biotic stress, its high abundance in the apoplast (5th highest intensity at 3.56% share of intensity in *S.lycopersicum*) and strong positive selection acting on three residues ($\omega = 8.1$ and $\rho = 0.1$) strongly implicates it in plant immunity (**Table 4**). Two surface-exposed residues under positive selection (position I and II) are close to the catalytic residues (**Figure 22**). Position I has either threonine or alanine and position II is closer to the active site encoding either serine or threonine. Neither substitution is drastic, but each alters the biochemical environment and could disrupt direct protein-protein interaction involving these residues. Since at least one site under positive selection is close to the active site, we consider Solyc01g105070.3.1 to show a weak 'ring of fire' as a potential footprint of an antagonistic molecular arms race (Bishop et al. 2000; Bishop et al. 2005).

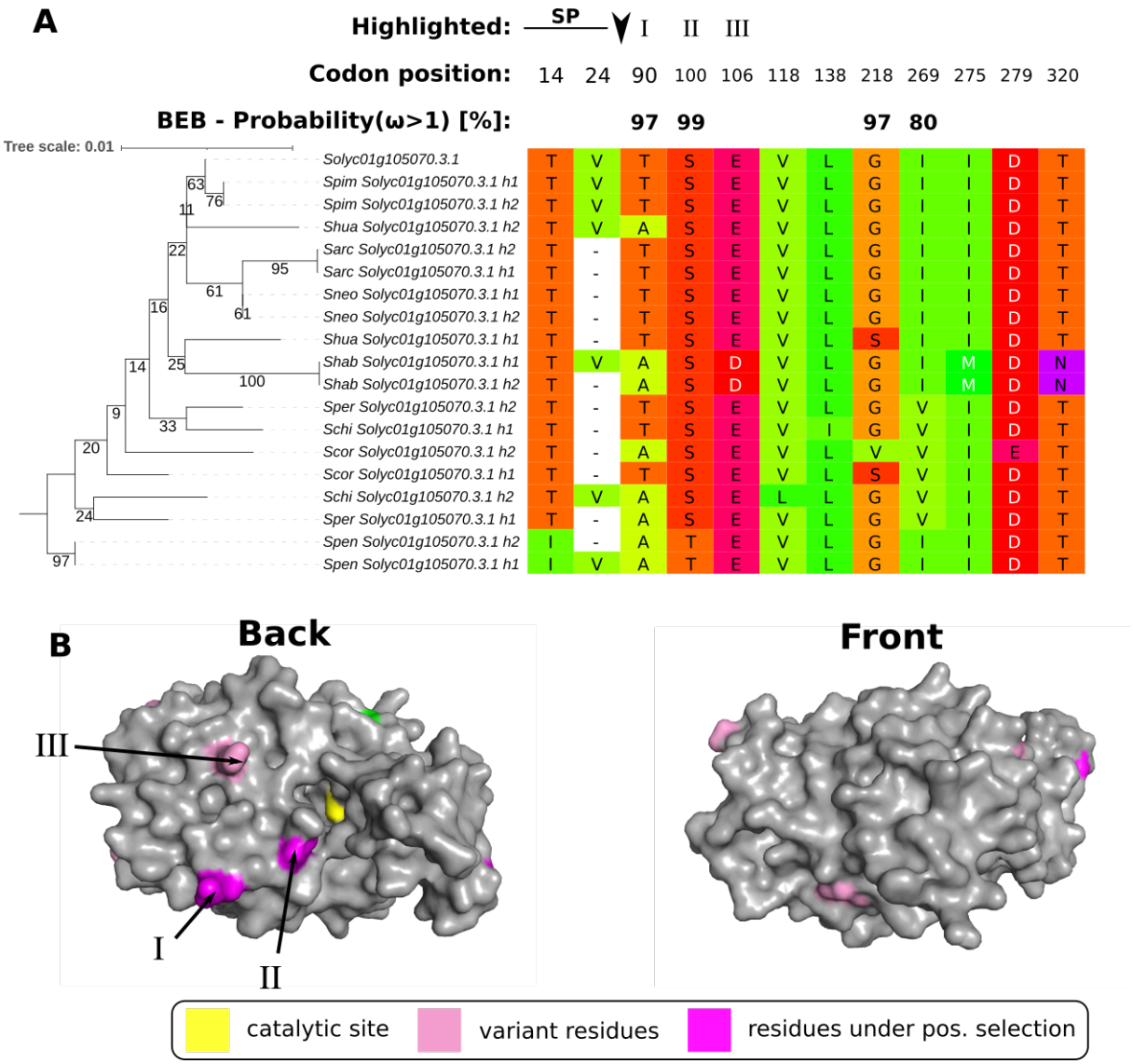


Figure 22 Combined proteomics and molecular evolution data implicates the haem peroxidase Solyc01g105070.3.1 in plant immunity. (**A**) Summary of positive selection analysis comprising the nucleotide-based maximum likelihood phylogenetic tree with bootstrap values and multi sequence alignments of all variant amino acid residues. SP indicates the SignalP predicted signal peptide and Bayes empirical Bayes (BEB) indicates the posterior probability of codon positions to be under positive selection. (**B**) 3D model of Solyc01g105070.3.1 based on the radish peroxidase template 4a5g.1.A. Residues identified under positive selection are highlighted in magenta.

2.4 Discussion

In this chapter, we extracted apoplastic fluid from three distantly-related tomato species treated with and without Bion. Then, we generated a reference proteome for *S.peruvianum* to analyse its AF extracts via mass spectrometry. Comparative proteomics addressed

our first question about similarities and differences in how different tomato species defend their apoplasts upon perceived biotic stress: Biotic stress-induced AFs were similar across tomato species, as approximately one third of proteins in Bion-induced proteomes belonged to the same five orthogroups. These five orthogroups were conserved in their biotic stress-response and highlighted clear similarities between apoplastic defences of different tomato species. Furthermore, we found that these five pillars of induced defence were supplemented by many Bion-induced orthogroups with low LFQ intensities. These orthogroups marked a stark diversity between tomato species in their apoplastic biotic stress responses. Finally, we analysed positive selection in Bion-induced apoplastic proteins and identified sites of positive selection in known immunity-related proteases Pip1 and P69B and used this approach to discover 19 additional tomato proteins that might act in apoplastic defences. The workflow to identify these 21 proteins is displayed in **Figure 23**.

2.4.1 Protein concentration in wild tomato apoplastic fluid extracts increases upon Bion treatment

We observed a stronger increase in apoplastic protein concentration upon Bion treatment in *S.peruvianum* and *S.pennellii* compared to *S.lycopersicum* (**Figure 6B**) and this increase was not caused by increased cytoplasmic contamination (**Figure 6**). This showed that upon biotic stress, wild tomato species fortified their extracellular space by increasing protein concentration in the apoplast. The small and statistically insignificant increase in abundance of apoplastic proteins in *S.lycopersicum* in contrast to wild tomatoes could be interpreted as *S.lycopersicum* mounting a sparser immune response upon biotic stress. This would be in line with a domestication process which might have favoured growth and yield over immunity; especially with the supportive exo-immune system in the form of pesticides in recent decades. Comparative genomics showing loss of defence-related genes during domestication of tomato support this hypothesis (Gao et al. 2019b). In all tomato species, however, we saw increasing activity of serine hydrolases of comparable size to P69 subtilases, which include the immunity-related P69B (**Figure 7**) (Sueldo et

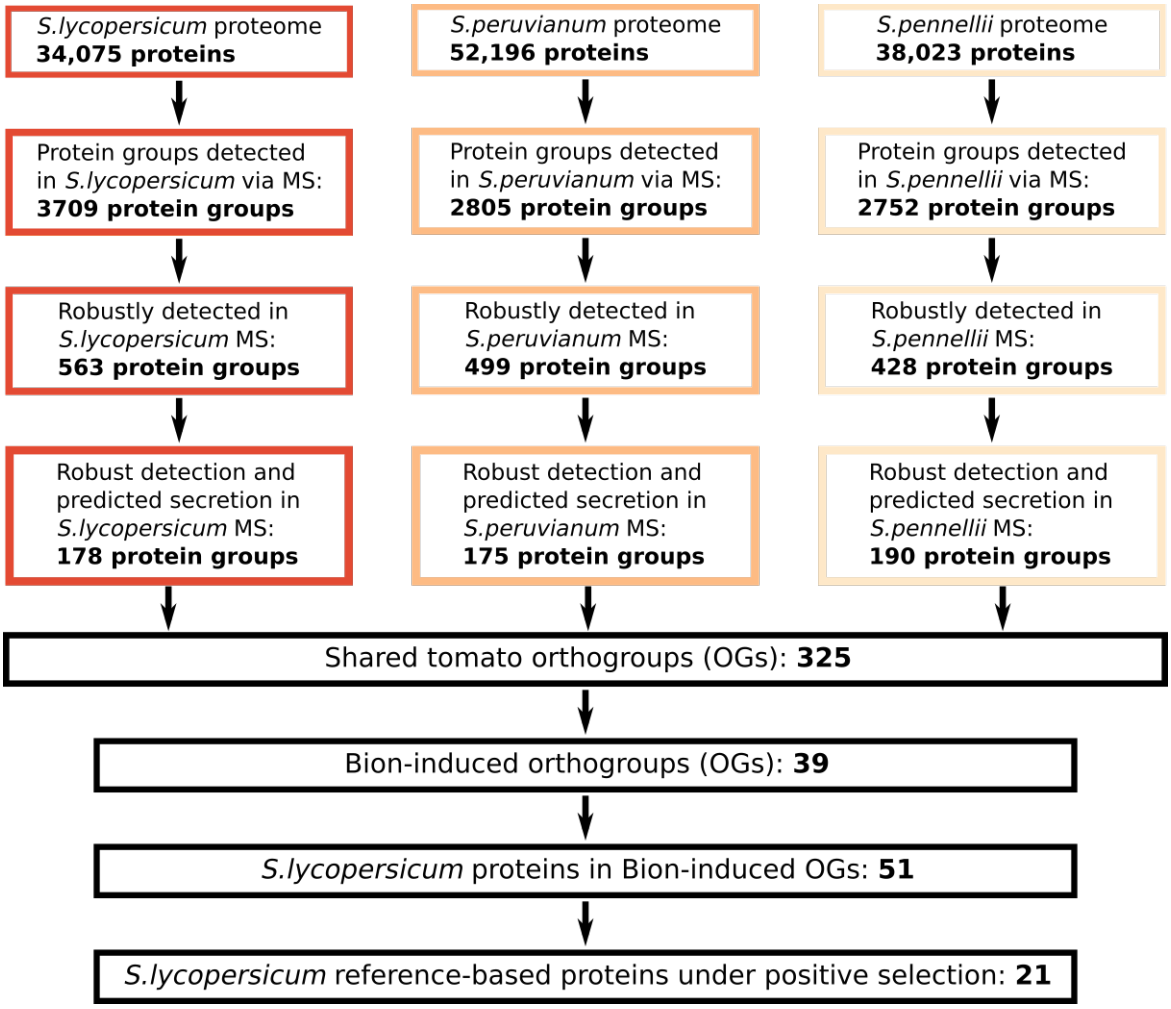


Figure 23 Flow chart describing the identification of 21 tomato proteins implicated in plant immunity.

al. 2014). This increase in serine hydrolase activity was strongest in *S.lycopersicum*. The strong increases of abundance and activity of P69 subtilases during biotic stress had been shown for different domesticated tomato cultivars and domesticated potato (*Solanum tuberosum*) before (Tian et al. 2004; Ali et al. 2014; Planas-Marquès et al. 2018). The increase of P69 activity occurs across different tomato species, making it likely that increase of P69 subtilase activity upon biotic stress is a conserved immunity feature across tomato species and not a domestication-induced property of *Solanum* species.

2.4.2 Assembling a genome and proteome of *Solanum peruvianum*

We generated a reference proteome of *S.peruvianum* to annotate MS spectra. This proteome had similar levels of completeness compared to published tomato proteomes based

on BUSCO comparisons (**Figure 9**), but there are shortcomings of this proteome. First, the generated proteome was likely too large, containing nearly 15,000 extra proteins compared to the currently largest published wild tomato proteome of *S.pennellii* (Bolger et al. 2014a). The proteome could have been reduced to high-confidence proteins by incorporating published RNAseq data. Second, this approach did not assemble genomic reads that were not matched to the *S.lycopersicum* reference. These unmapped reads could have contained genomic stretches absent from the *S.lycopersicum* genome assembly. For our purpose however, this proteome was sufficient.

2.4.3 Comparing secretomes between different tomato species

Orthogroup-based analysis of MS data is invaluable for comparative proteomics. Several issues made comparing different MS data sets difficult in the past. Prominent was the lack of a shared reference which could be overcome by using the identical proteome to analyse closely related individuals (Ali et al. 2014). While this might be feasible for individuals from the same species, information will be lost when analysing MS data from more distantly-related species. Via orthogroup-based analysis, we clustered proteins based on homology to generate a shared reference with little loss of information. Apoplastic orthogroups contained only slightly more than a single protein per orthogroup (1.1, 1.2 and 1.3 for *S.lycopersicum, S.peruvianum* and *S.pennellii* respectively), allowing for accurate and near one-to-one gene comparisons between the species. Furthermore, centering normalised intensities around zero preserved relative changes between treatments. Moreover, comparing shares of protein intensities of total sample intensities provided an assessment of the relative abundance of different orthogroups in their respective protein samples. Last, technical differences can have been introduced by different handling of samples and specifications of MS analyses tools. Since all samples were prepared by the same person and analysed by the same mass spectrometer, this bias will be small. We found 39 orthogroups that increasingly accumulate in the tomato apoplastic fluid upon Bion treatment. These orthogroups contained classes of proteins known to accumulate during biotic stress in domesticated *Solanum* species and included subtilases and papain-like cysteine

proteases, protease inhibitors, peroxidases, GDSL lipases, and glycosyl hydrolases such as chitinases, endo-1,3-β-glucanases and others. (Parker et al. 2013; Ali et al. 2014; Planas-Marquès et al. 2018; Abreha et al. 2021). Only five of 39 orthogroups increased in all three tomato species, highlighting few conserved apoplastic responses across different tomato species. The disproportionate dominance of these five orthogroups was demonstrated by their share of total intensities in each sample and we estimated that approximately every third protein in apoplastic fluid belonged to only these five orthogroups (**Figure 15B, C, D**). The orthogroup containing P69 subtilases stood out as the most abundant in all three tomato biotic stress-triggered AF extracts, accounting for 18.7-27.5% of total protein intensity. The stress-induction and dominance of P69 subtilases had been documented for domesticated *Solanum lycopersicum* and *Solanum tuberosum* (potato) before (Planas-Marquès et al. 2018; Abreha et al. 2021), but it was not known for wild *Solanum* species. In the more distantly related *Solanaceae* species *Nicotiana benthamiana* (wild tobacco), subtilases are also dominant in AF extracts, but they are not induced upon biotic stress signalling (Grosse-Holz et al. 2018). Our data showed that increased accumulation of P69 subtilases is a conserved feature among domesticated and wild *Solanum* species. Here, we focused on analysing orthogroups increasing in abundance upon Bion treatment. Future work could investigate orthogroups disappearing from AFs upon biotic stress signalling using this data set. This section of the chapter raised several questions. First: what are the biological functions of the 39 Bion-induced orthogroups? Deducing the biological function of a protein based on its functional annotation is challenging. For instance, four different orthogroups encoded chitinases which comprised two different PR protein groups (Sinha et al. 2014). The PR3 chitinases are again subdivided into different classes of chitinases (Sinha et al. 2014). Ongoing work in our lab has confirmed that a protein annotated as a PR3 chitinase from *Nicotiana* is indeed involved in immunity, but against bacteria, not fungi, and this PR3 does not hydrolyse chitin (Godson et al. *in preparation*). Another emerging question was how differences and similarities in tomato apoplastic proteome landscapes compare to other, more distantly-related plant species. For instance, will the same orthogroups emerge to be the core components of apoplastic

biotic stress responses? This question will be answered in the next chapter.

2.4.4 Detecting molecular evolution in biotic stress-induced tomato secretomes

Positive selection implicated 21 proteins in plant immunity by being both Bion-induced and under positive selection. Of those 21 proteins, P69B and Pip1 presented prominent positive controls, both with known immunity-related functions and high amino acid diversity. Here, we added statistical proof of them being under positive selection in tomato and analysed location and possible effects of residues under selection in more detail. Our approach implicates 19 additional tomato proteins in immunity.

2.4.4.1 P69B

P69B accumulates as a highly abundant protein in the apoplast upon biotic stress (Tornero et al. 1997; Planas-Marquès et al. 2018). So far, P69B has a dual role in plant immunity by activating the immunity-related host protease Rcr3 and by releasing an immunogenic peptide of a small secreted effector of *P.infestans* that elicits cell death (Paulus et al. 2020; Wang et al. 2021a). Release of the immunogenic peptide can be prevented by *P.infestans*-secreted inhibitors of P69B, the Kazal-like inhibitors Epi1 and Epi10 (Tian et al. 2004; Tian et al. 2005; Wang et al. 2021a). However, despite dominance of P69B in the apoplast and its dual role in immunity, no other pathogen-derived inhibitors have been described yet. Identifying additional, unrelated P69B inhibitors would support its role as a hub in tomato immunity. High amino acid variation between P69B alleles from *S.lycopersicum* and *S.pimpinellifolium* were described before, but positive selection had not been shown yet (Hörger and van der Hoorn 2013). Our analyses proved P69B to be under positive selection in tomatoes and identified several residues under selection that were close to the substrate binding groove. Especially position III showed highly variable amino acid residues that could disrupt interactions with cognate pathogen-derived inhibitors (**Figure 21**). Testing P69B variants in interactions with Epi1 and Epi10 are plausible next steps in characterising the residue's involvement in antagonistic arms races and activity-based

protein profiling (ABPP) with FP-TAMRA would be a suitable tool. Not only P69B, but also P69C, P69D and P69J subtilases are under positive selection (Reichardt et al. 2018). Almost nothing is known about the function of other P69 homologs. Some functional redundancy between P69 subtilases is imaginable, but they are known to show clear substrate specificity. For instance, only P69B and not the other P69 homologs were able to release the an immunogenic peptide from *P.infestans* effector PC2 (Wang et al. 2021a). Their roles remain to be explored.

2.4.4.2 Pip1

Pip1 (Solyc02g077040.4.1) accumulates upon biotic stress and is involved in basal plant defence against a broad range of pathogens (Tian et al. 2007; Shabab et al. 2008; Ilyas et al. 2015). Pip1 is targeted and inhibited by different, unrelated plant-pathogen inhibitors, such as EpiC2B of *P.infestans* and Avr2 of *C.fuvlum* (Tian et al. 2007; Shabab et al. 2008). Its increase in abundance upon biotic stress, contribution to basal defence and targetig by unrelated plant pathogen-derived inhibitors, highlight Pip1 as a potential hub in tomato apoplastic immunity. Pip1 accumulated disproportionately more non-synonymous mutations leading to biochemically dissimilar over similar amino acids in tomatoes (Shabab et al. 2008). However, statistical evidence of disproportionate accumulation of mutations resulting in non-synonymous amino acid substitutions compared to mutations resulting in synonymous amino acid substitutions had not been provided yet. Previously, Shabab et al. 2008 analysed sequenced PCR products, amplified from cDNA obtained from different wild tomato species which yielded eight different Pip1 alleles from eight tomato species. In contrast, by mapping genomic reads we identified 20 different Pip1 alleles from twelve tomato species which increased the resolution in our analysis. Both analyses identify codons 139 and 314 to encode different amino acids (**Figure 20**) (Shabab et al. 2008). While codon 139 encoded similar amino acids (arginine and lysine) before, our analysis identified an additional allele encoding tryptophan in *S.pennellii*. Codon 314 encodes different amino acids in both analyses and our data added glycine as another alternative amino acid to alleles known to encode aspartic acid and serine. An interesting difference

between the analyses was codon 238, which encodes asparagine and serine (Shabab et al. 2008). While our analysis identified the same alternative amino acids, codon 238 was not identified to be under positive selection. Inspection of the amino acid alignment coupled with the gene tree showed that asparagine and serine were species-clade specific variants (**Figure 20**). This more likely points towards a shared common ancestor with the alternative allele than a repeated accumulation of alternative amino acids due to ongoing positive selection pressure.

2.4.4.3 Solcy01g105070.3.1 - class III haem peroxidase

Class III peroxidases are recognised PR9 proteins (van Loon et al. 2006) and contribute to plant immunity by cell wall reinforcement through extensine linking, lignification and suberin formation and by generating ROS and phytoalexins (Almagro et al. 2009). The maize class III peroxidase POX12 increases resistance against the fungus *Ustilago maydis* (Hemetsberger et al. 2012) and is targeted by apoplastic Pep1 of *U.maydis* (Doehlemann et al. 2009). The *Zm*POX12-*Um*Pep1 system illustrates an arms race the class III peroxidases could be locked in. Furthermore, stable overexpression of a class III peroxidase in sweet oranges (*Citrus sinensis*) increased resistance to citrus bacterial canker, likely by contributing to elevated H_2O_2 levels and increased lignification (Li et al. 2020). Molecular evolution has not been analysed for either peroxidase yet, but evolutionary analyses of class III peroxidases have focused on comparison of all homologs within *A.thaliana*, maize (*Zea maize*), Chinese Pear (*Pyrus bretschneideri*) and silver birch (*Betula pendula*) (Kupriyanova et al. 2015; Wang et al. 2015; Cao et al. 2016; Cai et al. 2021). These studies, calculated Ka/Ks ratios by pairwise comparisons of class III peroxidases within species and found predominantly purifying selective pressures between homologs of the same species. These published analyses indicated little diversification within plant species. However, to avoid mistaking strong positive selection for purifying selection, analyses of dN/dS ratios are best conducted between closely-related species (Kryazhimskiy and Plotkin 2008).

Here we found positive selection pressure acting on a few surface exposed residues in proximity to the catalytic residues. Solyc01g105070.3.1 might be involved in plant immunity either by contributing to a ROS burst, cross-linking to reinforce the cell wall or involved in the generation of phytoalexins. Especially the extensine-like LRR (Solyc02g064940.1.1) could be an interesting target for oxidation by Solyc01g105070.3.1.

Central open questions emerge from this chapter: Can we prove a role for these 21 proteins in plant immunity? Are these residues under positive selection imprinted by antagonistic arms races and if so, which cognate proteins are causing the selection pressure? The first question can be addressed by evaluating these proteins in plant infection assays. Transient heterologous over expression in *N.benthamiana* can be central to describing tomato plant immunity hydrolases and recently, high-throughput tools have been generated in our lab (Paulus et al. 2020; Kourelis et al. 2020; Buscaill et al. 2021; Wang et al. 2021a; Jutras et al. 2021). To test if positive selection might be due to an arms race, variant proteins should be tested in inhibition assays with known inhibitors. Last, we do not know how rates of positive selection in genes encoding Bion-induced proteins compare to genes encoding non-induced proteins. It will be interesting to compare rates of positive selection between genes encoding Bion-induced and non-induced proteins in future.

3 Immuno Secretomes are different between plant taxons

3.1 Introduction

In the struggle for limited resources, plant pathogens attempt to colonise and replicate in plant hosts which in turn defend themselves fiercely against attempted invasions. One such molecular battlefield is the intercellular space, the apoplast, where results of plant-pathogen interactions can be decisive for the outcome of attempted infections (Doehlemann and Hemetsberger 2013). Plants thus fortify their apoplasts upon perception of diverse infection-associated triggers by secreting defence-related proteins. Some defence-related proteins, increasing drastically in abundance after perception of infection-associated triggers, are referred to as pathogenesis-related (PR) proteins (van Loon et al. 2006). Highly-abundant PR proteins form part of the environment a pathogen has to navigate and adapt to in order to successfully colonise the host. The entirety of host proteins defining the apoplastic environment during plant defence is what we coined as the immuno secretome (IS). Immuno secretomes differ between distantly-related plant species and specialisation of pathogens to particular immuno secretomes likely reduce their ability to colonise other plant species. Different composition of immuno secretomes of distantly-related species thus likely play a role in non-host resistance (NHR) (recent review on NHR (Panstruga and Moscou 2020)). In resistance breeding, this has been taken advantage of by identifying apoplastic antimicrobial proteins in one species to over express them in another species (reviewed in Stintzi et al. 1993).

In the past years, proteomics has emerged as a powerful tool to study composition of immuno secretomes. To understand the entire composition of apoplastic fluid (AF) extracts, whole-sample shotgun proteomics has been increasingly deployed recently. Studies included the pathosystems rice-*Magnaporthe oryzae*, rice-*Cochliobolus miyabeanus*, wheat-*Zymoseptoria tritici*, tomato-*Ralstonia solanacearum* and *Nicotiana benthamiana*-*Agrobacterium tumefaciens* (Kim et al. 2013; Kim et al. 2014a; Yang et al. 2015; Planas-

Marquès et al. 2018; Grosse-Holz et al. 2018). These projects aided identification of proteins with crucial roles in the selected pathosystems, but made comparative proteomics difficult. Different pathogens likely constituted different biotic stress triggers as shotgun proteomic studies included different hemibiotrophic and necrotrophic fungi and different biotrophic bacteria. In addition to different biotic stress triggers, different sample preparation protocols and mass spectrometry methods likely introduced different technical biases into these data sets. Thus, recent advances in shotgun proteomics of biotic stress-induced apoplastic fluid extracts have progressed the understanding of individual pathosystems, but no holistic comparative proteomics project has been undertaken. Fundamentally, we do not have a global understanding of similarities and differences between immuno secretomes of different plant species.

In this chapter, we aim to understand similarities and differences between immuno secretomes of different plant species to map the apoplastic environments plant pathogens have to navigate and adapt to when attempting to infect different hosts. We will induce biotic stress in different plant species with the same trigger, the functional salicylic acid analogue BTH, the active ingredient of the product Bion. We will extract apoplastic fluid, analyse proteomes by MS and compare apoplastic proteomes between different species based on orthogroups. Then we will compare apoplastic proteomes using various metrics such as number of different orthogroups, principle component analysis, PR protein accumulation and abundance of functional groups.

3.2 Materials and Methods

3.2.1 Plant growth conditions

All plants were grown in soil and under conditions described in **Table 6**.

Table 5: Plant growth conditions

Species	Accession/Cultivar	Plant age [weeks]	T [°C]	Photop.* [hrs]	location
Solanum tuberosum	Maris Piper	6-7	21	16	Greenhouse
Nicotiana benthamiana	Lab strain	4	22-24	16	Greenhouse
Glycine soja	pi366124 Kagawa	-	21	8	Growth cabinet
Glycine max	Williams 82	4	21	16	Phytochamber
Phaseolus vulgaris	Tendergreen	4	20-22	8	Greenhouse
Arabidopsis thaliana	Col-0	5	12	20	Growth cabinet
Chenopodium quinoa	Atlas	4	22-24	16	Greenhouse
Mirabilis jalapa	Chiltern Seeds Nb. 880	6	22-24	16	Greenhouse
Hordeum vulgare	golden promise	4-5	22-24	16	Greenhouse
Marchantia polymorpha	9-11	Tak1 & Tak2		Grown by Hugh Mulvey	

*Photop.: Photoperiod

3.2.2 Apoplastic Fluid extraction

To extract apoplastic fluid, we syringe-infiltrated leaves of four week old *Nicotiana benthamiana* LAB and *Glycine max* cultivar Williams 82 and five week old *Arabidopsis thaliana* Col-0 with 100 mM phosphate buffer (pH 6.5) using a disposable 50 ml syringe. The choice of phosphate buffer was based on its comparably low level of extracted cytoplasmic contaminants (Witzel et al. 2011). Each biological replicate contained three (*N.benthamiana*) or four (*G.max*) leaves from the same plant. For *A.thaliana*, each biological replicate contained non-necrotic and non-senescent leaves from two to three different plants. After infiltrating leaves, we dried their surfaces by placing them between white fibre roll paper and gently applying pressure to remove surface water before rolling them up carefully to fit into disposable 20 ml syringes. Each syringe was placed into a sterile 50 ml falcon tube and stored on ice until all samples were ready for centrifugation. Next, we extracted apoplastic fluid from infiltrated leaves by centrifugation for 25 minutes, at 4°C and 2000 x g for *A.thaliana* and *G.max* and 500 x g for *N.benthamiana*. After extraction, apoplastic fluid was aliquoted and flash frozen in liquid nitrogen as quickly as possible. Leaves had been weighed before syringe-infiltration and the amount of extracted apoplastic fluid was measured to calculate the extracted amount per fresh weight. Total extracts were made from leaf discs taken from the same plants used for apoplastic fluid extraction.

Leaf discs were weighed and flash frozen in liquid nitrogen in 1.5 ml Eppendorf tubes with 2 metal beads. Leaf tissue was pulverised with a tissue lyser at a frequency of 30 per seconds for 1 minute. The tissue lyser blocks had been pre-cooled at −80°C beforehand. Pulverised tissue was mixed with 200 µl of ice-cold PBS buffer (0.137 M NaCl, 0.0027 M KCl, 0.01 M Na_2HPO_4 and 0.0018 M KH_2PO_4) on ice and leaf debris removed by centrifugation at 13,000 x g and 4°C for ten minutes. Supernatant was separated from the leaf debris and stored on ice to be used directly or flash frozen in liquid nitrogen.

3.2.3 Generating an unrooted species phylogenetic tree from proteomes

Proteomes from the NCBI RefSeq *A.thaliana* Tair10.1 proteome, *G.max* RefSeq proteome GCF_000004515.5 and the *N.benthamiana* NbDE were clustered in orthogroups based on sequence similarity using Orthofinder2.0 (Emms and Kelly 2019). We used Orthofinder2.0 to infer the underlying species tree from whole proteomes. In orthofinder, we used dendroblast and fasttree 2 to infer approximately-maximum likelihood phylogenetic trees (Kelly and Maini 2013; Price et al. 2010). The species tree was inferred by the Species Tree Inference from All Genes (STAG) method using closely-related single- and multi-copy orthogroups. The entire workflow to infer orthogroups and the unrooted species tree is encoded in the Orthofinder2.0 command line tool.

3.2.4 Generating a coding sequences-based phylogenetic tree

To generate a phylogenetic tree for 70 subtilases, we used coding sequences (CDSs) from *A.thaliana* Tair10.1 NCBI RefSeq (GCF_000001735.4), *G.max* v2.1 NCBI RefSeq (GC F_000004515.5), *S.pennellii* NCBI RefSeq (GCF_001406875.1), *S.lycopersicum* ITAG4.0 CDS, NbDE transcriptome and our own *S.peruvianum* CDS file from **chapter 2**. Coding sequences were aligned using mafft v.7.490 with the global alignment setting and a maximum of 1000 iterations (Katoh and Standley 2013). Codon multi sequence alignments were used to reconstruct a maximum-likelihood phylogenetic tree using RAxML v8.2.12 with a GTR substitution matrix and a Gamma model of rate heterogeneity and the seed 123 (Stamatakis 2014). RAxML was run on the University of Oxford Advanced Research

Computing (ARC) facility (Richards 2015). The phylogenetic tree was visualised using iTOL (Letunic and Bork 2021).

3.2.5 Enzymatic characterisation of apoplastic fluid extracts

Protein concentrations and cytoplasmic contamination of apoplastic fluid and total extract samples were measured as described in **section 2.2.4**. In this chapter, we diluted total extract samples of *G.max* and *N.benthamiana* 1:100 and 1:1000 for *A.thaliana*. Average volume of extracted apoplastic fluid was based on a determined average per g of fresh leaf in the respective experiments.

3.2.6 Preparing protein samples for mass spectrometry analysis

Samples were prepared using a filter aided sample preparation (FASP) protocol established in our lab by Dr Kyoko Morimoto. As filters, we used Vivacon 500 (10,000 MWCO) and for each sample, we prepared a total amount of 50 µg of protein. Filters were prepared by adding 200 µl of 0.1% trifluoroacetic Acid (TFA) in 50% acetonitrile and centrifuged at 15,000 x g for 15 minutes at RT. Next, 200 µl of 8 M urea in 100 mM triethylammonium bicarbonate buffer (TEAB) were added with 50 µg of protein sample and centrifuged at 15,000 x g for 15 minutes followed by four washing steps with 8 M urea in 100 mM TEAB. Next, proteins were denatured by again adding 8 M urea in 100 mM TEAB in addition to 4 µl of 500 mM (tris 2-carboxyethyl) phosphine hydrochloride (TCEP HCl), 20 µl of 500 mM 2-chloroacetamide, incubated in the dark at RT for 30 minutes and centrifuged at 15,000 x g for 15 minutes at RT. The filter was washed twice with 200 µl 6 M urea in 50 mM TEAB and centrifuged at 15,000 x g for 15 minutes at RT. Next Lys-C solution was added at a ratio of 5 µg of Lys-C per 200 µg of initially applied protein. Lys-C had been dissolved in 1 M urea in 50 mM TEAB at a concentration of 5 µg/100 µl. After over-night incubation at 37°C, we added trypsin at a ratio of again 5 µg trypsin per 200 µg of initially applied protein and incubated for 6-8 hours at 37°C. Trypsin had been dissolved in 100 µl 50 mM TEAB at a concentration of 5 µg/100 µl. Finally, filters were centrifuged at 15,000 x g for 15 minutes at RT, washed with 150 µl 0.1% TFA, then 150 µl

0.1% TFA in 50% acetonitrile and the flowthrough stored in a protein low-binding tube. Samples were not dried but sent on dry ice to our collaborator Dr. Farnusch Kaschani at the Analytics Core Facility Essen, Germany.

3.2.7 Mass spectrometry data processing and statistical analyses

Protein MS data was generated by our collaborator Dr Farnusch Kaschani. Peptides were identified using default settings in MaxQuant v1.6.10.43 using Andromeda as a search engine with label-free quantification (LFQ) and match between runs turned on (Cox and Mann 2008; Cox et al. 2011; Cox et al. 2014). All LFQ values are based on MaxQuant's MaxLFQ normalisation method (Cox et al. 2014). Static modification included carbamidomethyl and dynamic modifications included N-terminal acetylation, oxidation and deamidation. Oxidation was a modification included in quantification. Searches were performed against the NCBI RefSeq *A.thaliana* Tair10.1 proteome, *G.max* RefSeq proteome GCF_000004515.5 and the *N.benthamiana* NbDE reference proteome, respective reverse databases and a database of known contaminants. To benefit from identifications from total leaf extract protein samples in identifying peptides in apoplastic fluid samples without distorting values during normalisation, we analysed apoplastic fluid and total extract samples as separate parameter groups with a minimal LFQ ratio count of 2.

3.2.8 Analysis of protein MS data of apoplastic fluid extracts

All filtering steps were the same for each species with their cognate predicted secretomes and analysis were done in Jupyter-Notebooks (Kluyver et al. 2016; Homma 2022d). Since tryptic digest of different proteins can result in identical peptides, MS analyses can not always distinguish between different proteins. Groups of proteins that match identified peptides and for which insufficient evidence was collected to confidently identify the individual proteins are thus grouped into protein groups. These protein groups mostly contain proteins with high sequence similarity. First, protein groups were filtered by removing groups identified as contaminants, identified by matching only and labelled as reverse.

Reverse proteins derived from the decoy search database (reversed reference proteome) are false positives. Next, only protein groups identified in at least three out of four replicates in either or both treatments were considered to be detected robustly. No values were imputed. PCA analyses were done with centred protein groups implemented in sklearn (Pedregosa et al. 2011). Log2-transformed LFQ intensities were centred per protein group by subtracting the protein group's mean from each sample. Orthogroups were generated for whole proteomes of *S.lycopersicum*, *S.pennellii*, our generated *S.peruvianum* proteome (**chapter 2**), *N.benthamiana*, *A.thaliana* and *G.max* using Orthofinder v2.0 (Emms and Kelly 2019). Intensities for each orthogroup and species were determined by the sum of untransformed LFQ intensities of each contributing protein. Log2-transformed, summed-up LFQ intensities were centred per orthogroup and species by subtracting the orthogroup's and species' mean from each respective sample. To increase power when detecting proteins increasing in abundance upon Bion treatment, we selected a minimal log2FC threshold of 0.9 for proteins to be considered for statistical analysis of differential abundance. This value was based on a preliminary MS data exploration of known tomato PR7 which visibly accumulates on protein gels and is known to increase in abundance upon Bion treatment (Tian et al. 2004). Changes were calculated using unpaired, two-sided pairwise t-tests. For each species, we corrected p-values of all pairwise comparisons for multiple testing using Benjamini-Hochberg procedure. We used a significance threshold of 0.05 for multiple testing corrected p-values (sometimes referred to as q-values). To derive a proxy for protein concentration, we calcultated percentage shares of total intensities of orthogropus by dividing the untransformed LFQ intensity of an orthogroup by the sum of untransformed LFQ intensities of the respective sample. Functional annotations of proteins and orthogroups were based on NCBI annotations and InterPro Scan v5.50.84.0 Pfam codes (Jones et al. 2014). The python code of all analyses can be found in the GitHub repository (Homma 2022d).

3.3 Results

The aim of this chapter was to characterise similarities and differences between immuno secretomes of different plant species using a three-step approach: First, we screened ten different plant species for induction of biotic stress signalling via Bion treatment and feasibility of apoplastic fluid extraction. Second, we extracted biotic stress-induced and unstressed AF for *N.benthamiana, G.max* and *A.thaliana* and analysed their apoplastic proteomes by proteomics. Third, we compared Bion-induced apoplastic proteomes with an orthogroup-based approach, to identify and discuss key similarities and differences between apoplastic proteomes of different plant species.

3.3.1 Extracting biotic stress-induced apoplastic fluid extracts of different eudicot species

To study immuno secretomes in a broad range of plant species, we first tested conditions for growth-, biotic stress induction- and apoplastic fluid extraction for a broad range of plant species of either agricultural or scientific importance (**Table 6**). To induce biotic

Table 6: Species tested for AF extraction

Species	Name	Accession/Cultivar	Result
Solanum tuberosum	potato	Maris Piper	Fail - stressed control
Nicotiana benthamiana	wild tobacco	Lab strain	**Success**
Glycine soja	wild soybean	pi366124 Kagawa	Fail - growth[1]
Glycine max	soybean	Williams 82	**Success**
Phaseolus vulgaris	french bean	Tendergreen	Fail - stressed control
Arabidopsis thaliana	arabidopsis	Col-0	**Success**
Chenopodium quinoa	quinoa	Atlas	Fail - stressed control
Mirabilis jalapa	four o'clock flower	Chiltern Seeds Nb. 880	Fail - stressed control
Hordeum vulgare	barley	golden promise	Fail - stressed control
Marchantia polymorpha	liverwort	Tak1 & Tak2	Fail - AF extraction

1: Only six seeds were available and plants died before flowering

stress, we drenched soil daily for three days, each time with 10 ml water or 10 ml of 1 mM BTH. Solutions containing 1 mM BTH were generated by dissolving Syngenta's pesticide Bion, containing 50% (w/w) BTH, in water. AF was extracted six days after the first treatment. Clear accumulation of pathogenesis-related (PR) proteins and absence of PR protein accumulation in the mock controls were the criteria to assess AFs for biotic stress

induction and unstressed control plants respectively. We succeeded in extracting Bion-induced and unstressed AF for *N.benthamiana, G.max* and *A.thaliana*. For *S.tuberosum, P.vulgaris, C.quinoa, M.jalapa* and *H.vulgare*, mock-controls were stressed and consistently displayed PR protein accumulation. For *Glycine soja*, we failed to run experiments as we could not grow healthy plants to produce seeds from the few seeds available. For *Marchantia polymorpha*, syringe-infiltration and AF extraction did not yield detectable protein in apoplastic fluid extracts.

To grasp the genetic distance between the three eudicot species *N.benthamiana, G.max* and *A.thaliana* relative to the three tomato species *Solanum lycopersicum, peruvianum* and *pennellii* analysed in **chapter 2**, we generated an unrooted species phylogenetic tree, inferred from whole proteomes by Orthofinder 2.0 (Emms and Kelly 2019). In the tree, tomato species clustered together and the more distantly-related *N.benthamiana* formed a sister clade to tomatoes consistent with the four species belonging to the Solanaceae family (**Figure 24**). The rosids *G.max* and *A.thaliana* formed another distinct clade. The

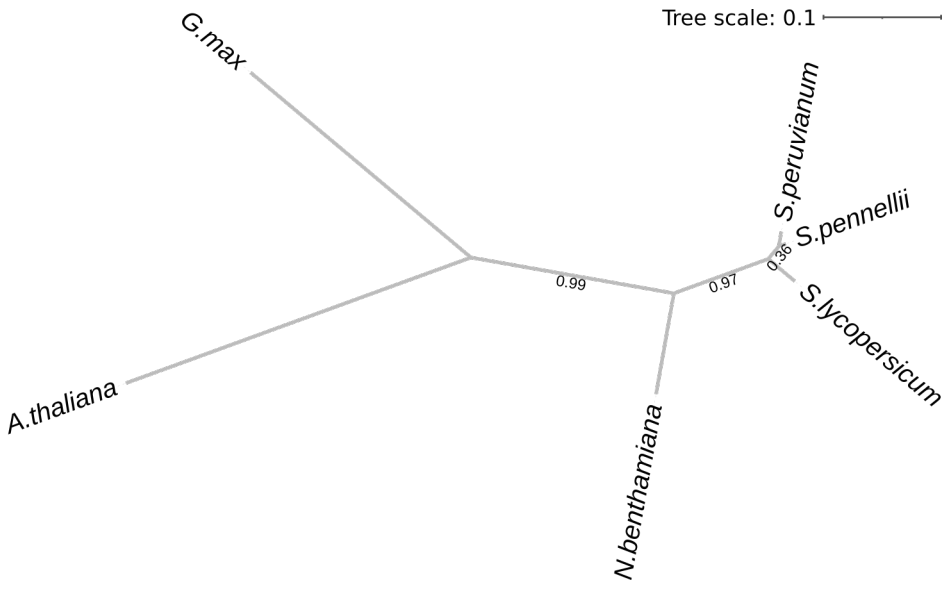


Figure 24: Unrooted, species phylogenetic tree of tomato species and three distantly-related eudicot species. A species phylogenetic tree inferred from whole proteomes of three tomato species (*S.lycopersicum, pennellii* and *peruvianum*) and *N.benthamiana, G.max* and *A.thaliana* by Orthofinder 2.0 (Emms and Kelly 2019). All bipartition support values are indicated.

profile of proteins accumulating in AFs is different between *N.benthamiana, G.max* and *A.thaliana* and contains different patterns of PR protein accumulation (**Figure 25A**).

This was in contrast to more similar profiles of PR proteins observed for tomato species in **chapter 2** (**Figure 6A**). For all three species, protein concentration of AFs changed

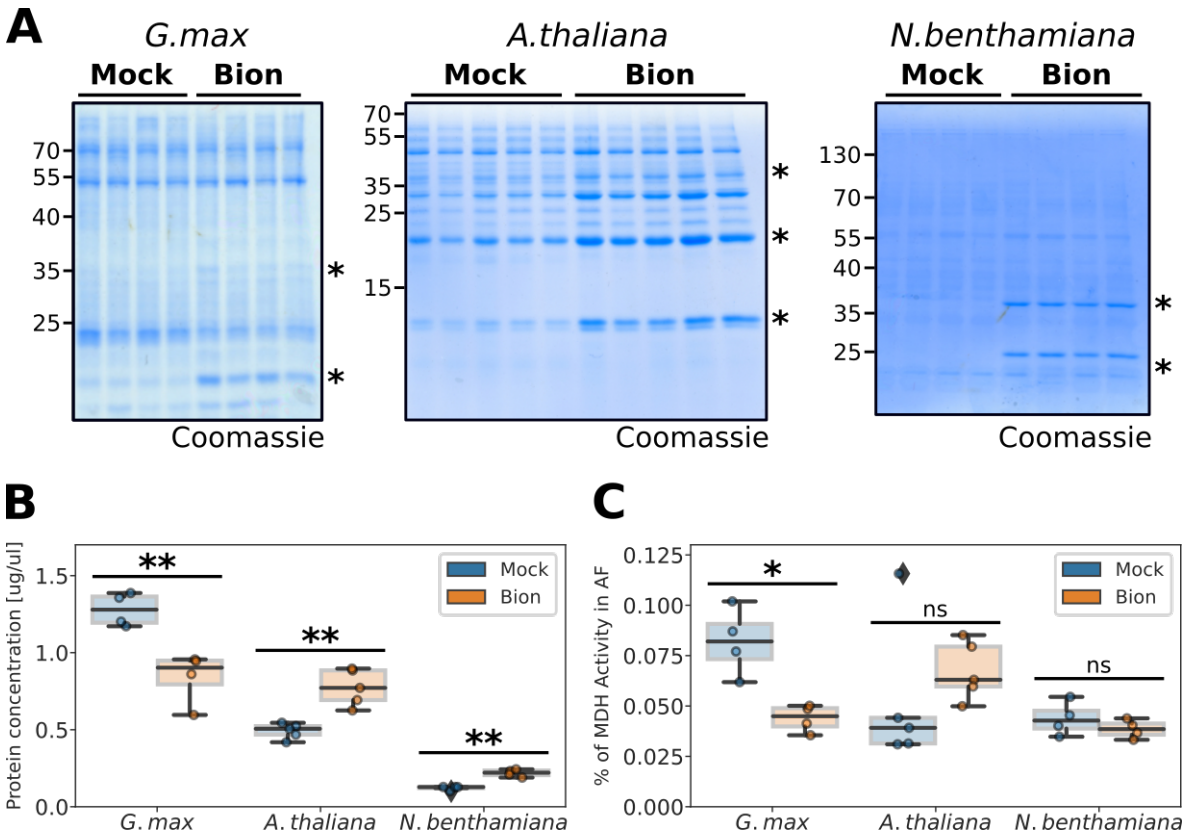


Figure 25: Extracting biotic stress-induced apoplastic fluid from different eudicot species. (**A, B, C**) All apoplastic fluid samples are aliquots of the same biological replicates from the same experiments. (**A**) Asterisks indicate accumulating PR proteins. For *N.benthamiana*, the gel was only resolved until shortly below the 25 kDa band. (**B**) Protein concentration of AF extracts from mock- and Bion-treated plants. (**C**) Cytoplasmic contamination measured as MDH activity in AFs relative to MDH activity in total extracts. Centre lines are the median, bounds of boxes the 25th and 75th percentiles and whiskers 1.5 x interquartile range (IQR) from 25th and 75th percentiles. Diamonds indicate outliers. Benjamini-Hochberg, false discovery rate-corrected p-values of two-sided, unpaired t-tests are indicated as: $*$: <0.05, $**$: <0.01.

upon Bion treatment (**Figure 25B**). For *A.thaliana* and *N.benthamiana*, protein concentration increased by 57.3% and 80.2% upon Bion treatment respectively. To put these values into context, protein concentration in AF of wild tomato also increased by 52% and 51% upon Bion treatment (**Figure 6B**). Surprisingly, protein concentration of *G.max* AF extracts decreased by 34.4% upon Bion treatment (**Figure 25B**). We did not expect this to be based on mislabelling of samples since AFs of Bion-treated plants showed a clear accumulation of PR proteins (**Figure 25A**). It might be correct, as one abundant 25

kDa protein accumulated less upon Bion treatment (**Figure 25A**). Malate-dehydrogenase (MDH) activities in AFs were between 0.03% and 0.12% of total extracts, indicating low cytoplasmic contamination in our AFs. For *N.benthamiana* and *A.thaliana*, we did not detect significant differences in MDH activity between AFs of mock- and Bion-treated plants (**Figure 25C**). For *G.max*, MDH activity was lower in Bion- than in mock samples (**Figure 25C**).

3.3.2 Preparing MS data for comparative proteomics

We next generated protein mass spectrometry data of apoplastic fluid extracts from mock- and Bion-treated plants for *N.benthamiana*, *G.max* and *A.thaliana*, comprising four biological replicates per treatment each. Overall, all log2-transformed protein groups ap-

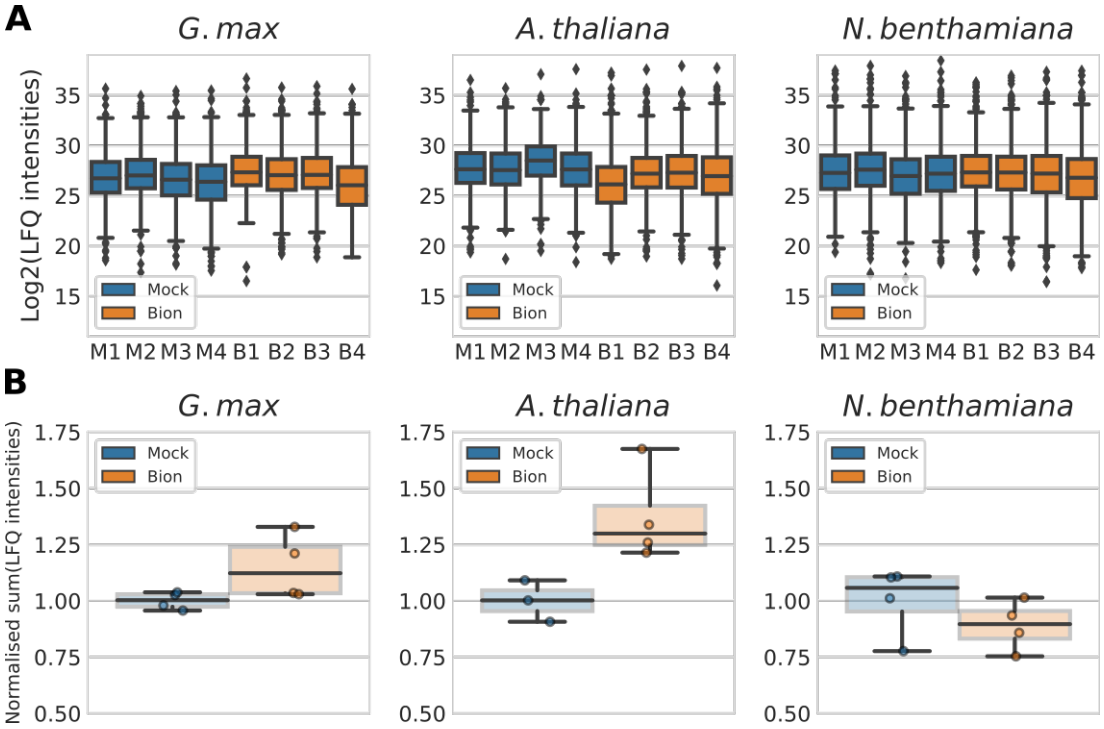


Figure 26: MS data of AF of soybean, *A.thaliana* and *N.benthamiana*. (**A**) Distribution of log2-transformed LFQ-normalised values for all protein groups detected in mock- (M1-4) and Bion-induced (B1-4) AFs of soybean (*G.max*), *A.thaliana* and wild tobacco (*N.benthamiana*). (**B**) Sum of LFQ-normalised values for each sample and the three species. For each species, sample intensities were normalised relative to the mean of the sum of LFQ-normalised mock-treated samples. Centre lines are the median, bounds of boxes the 25th and 75th percentiles and whiskers 1.5 x interquartile range (IQR) from 25th and 75th percentiles. Diamonds indicate outliers.

peared to be normally distributed and ranges of log2-transformed LFQ intensities ap-

peared similar between samples of the same species (**Figure 26A**). Adding up the non-transformed LFQ intensities for each treatment and species showed higher total LFQ intensities for *G.max* and *A.thaliana* (**Figure 26B**). For *N.benthamiana*, total LFQ intensity was lower in Bion-induced AFs compared to mock-treated AFs (**Figure 26B**) which was in contrast to Bion-induced protein concentration in *N.benthamiana* AFs. Next, we retained protein groups that were detected in at least three of four biological replicates of either or both mock- or Bion-induced samples. These protein groups were considered to be detected robustly. Of these, we only retained protein groups containing at least one protein predicted by either SignalP5.0, TargetP2.0 or ApoplastP1.0.1 to be secreted or apoplastic. Next, we centred values for each protein group around a mean of zero by subtracting the protein group's mean from the protein group's value in each respective sample. Centring values around a mean of zero maintained relative changes in abundance for protein groups, which is not the case for the widely-used z-score. To understand differences between samples of the same species, we subjected centred data sets to linear dimensionality reduction via principal component analysis (PCA). PCA of each species showed that most of the difference between samples was explained by Bion treatment (**Figure 27**). For *A.thaliana* however, PC2 explained more difference that

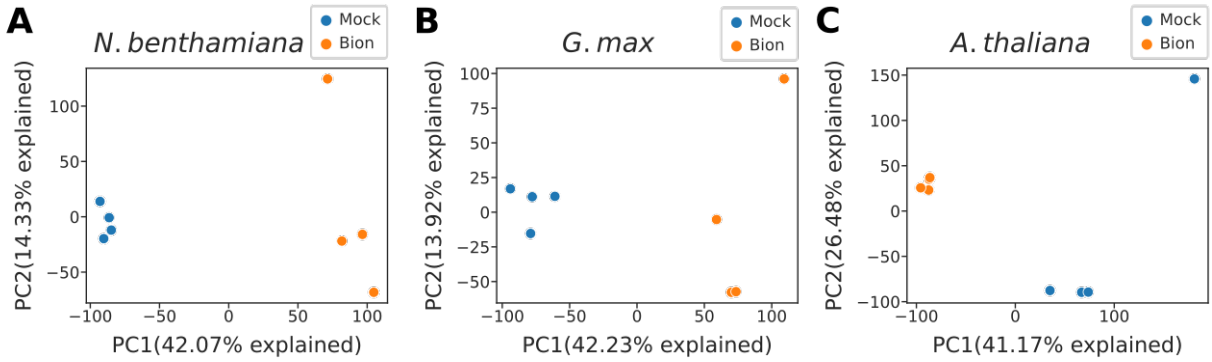


Figure 27 Bion-mediated biotic stress induction explains differences between MS samples. Principal component analysis or robustly detected, predicted to be secreted protein groups of *N.benthamiana* (**A**), *G.max* (**B**) and *A.thaliana* (**C**).

was not related to Bion treatment than for other species (26.48% relative to 14.33% for *N.benthamiana* and 13.92% for *G.max*) (**Figure 27**). Especially one mock sample appeared to be very different from other samples. We thus decided to have a closer look at

protein group distribution of *A.thaliana* samples and noticed that mock sample 3 did not contain lower-abundance protein groups shared by most other samples (**Figure 28A**). This mock sample 3 also had a relatively high MDH activity (**Figure 25C**). We therefore

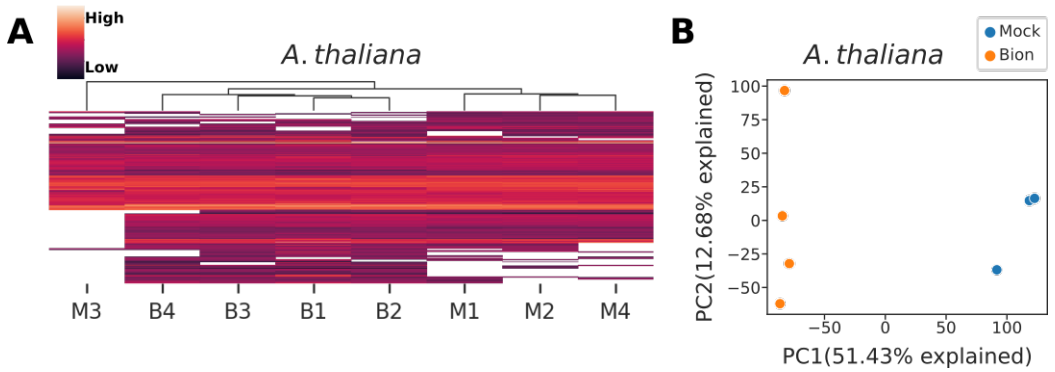


Figure 28 Removing a biological replicate for *A.thaliana* improves data quality. (**A**) Clustermap of biological replicates with robustly detected protein groups *A.thaliana* MS data. Data of mock- (M1-4) and Bion-treated (B1-4) samples had been log2-transformed and not centred. The clustermap was generated with a euclidean distance metric in combination with an averaging linkage method. (**B**) Principal component analysis of centred *A.thaliana* MS data without mock-treated sample 3.

decided to remove mock-treated sample 3 from all analyses without changing any of the downstream statistical procedures. Repeating *A.thaliana* PCA without mock sample 3, showed that more difference than before could be explained by Bion treatment and less within-treatment difference was explained by PC2 (**Figure 28B**).

Next, we aimed to understand basic similarities and differences between MS data of all six species. Apoplastic proteomes of *G.max, A.thaliana* and *N.benthamiana* contained 685, 975 and 805 robustly detected and predicted to be secreted protein groups, respectively. This is more than were detected for *S.lycopersicum* (563), *S.peruvianum* (499) and *S.pennellii* (428). To enable comparisons of apoplastic proteomes between different species, we clustered whole proteomes of *S.lycopersicum, S.peruvianum, S.pennellii, N.benthamiana, G.max* and *A.thaliana* into orthogroups (OGs) using Orthofinder2.0 (Emms and Kelly 2019). We detected 1.16±0.66, 1.41±1.41, 1.44±1.17, 1.95±1.29, 1.93±1.39 and 1.76±1.20 proteins per orthogroup for *S.lycopersicum, S.peruvianum, S.pennellii, N.benthamiana, G.max* and *A.thaliana*, respectively, meaning that our orthogroup-based approach mostly compares one to two proteins per species and orthogroup. This is a

lower resolution than a 1-to-1 comparison, but enables between-species comparative proteomics.

As in **chapter 2**, we calculated the intensity of an orthogroup by adding up untransformed, non-centred LFQ intensities of each contributing, robustly detected and predicted to be secreted protein group. We then standardised log2-transformed intensities of orthogroups to be centred around zero for each species. These centred log2-transformed intensities preserved the relative differences in intensities of each orthogroup within a species and could be used to directly compare different species to each other. To estimate abundance of orthogroups, we separately expressed each orthogroup's untransformed LFQ intensity as its share of the total untransformed LFQ intensity of each sample.

3.3.3 Comparative proteomics of six plant immuno secretomes

Across apoplastic proteomes of the six different plant species, we detected proteins from a total of 745 different orthogroups. In tomato species, we detected the lowest number of proteins from different orthogroups, with proteins from 162 orthogroups detected for both *S.lycopersicum* and *S.peruvianum* and proteins from 179 orthogroups detected for *S.pennellii* (**Figure 29**). By contrast, *N.benthamiana* AF contained proteins from 336 orthogroups, which is approximately twice as many as for tomato species. Furthermore, apoplastic fluid extracts of *G.max* contained proteins from 223 different orthogroups and apoplastic fluid extracts of *A.thaliana* contained proteins from 316 orthogroups. These results do not identify a pattern described either by evolutionary relationship, domestication status, quality of genome assembly or apoplastic fluid extraction force that explained the highly different numbers of orthogroups with proteomics evidence for each species.

3.3.3.1 Bion-induced changes in immuno secretomes

To explain differences between species and treatments, we ran a PCA analysis on centred orthogroup intensities of all 48 samples. Biological replicates of the same species and treatment clustered close together (**Figure 30**), but combining different species for PCA analysis results in less difference explained by Bion treatment compared to single-species

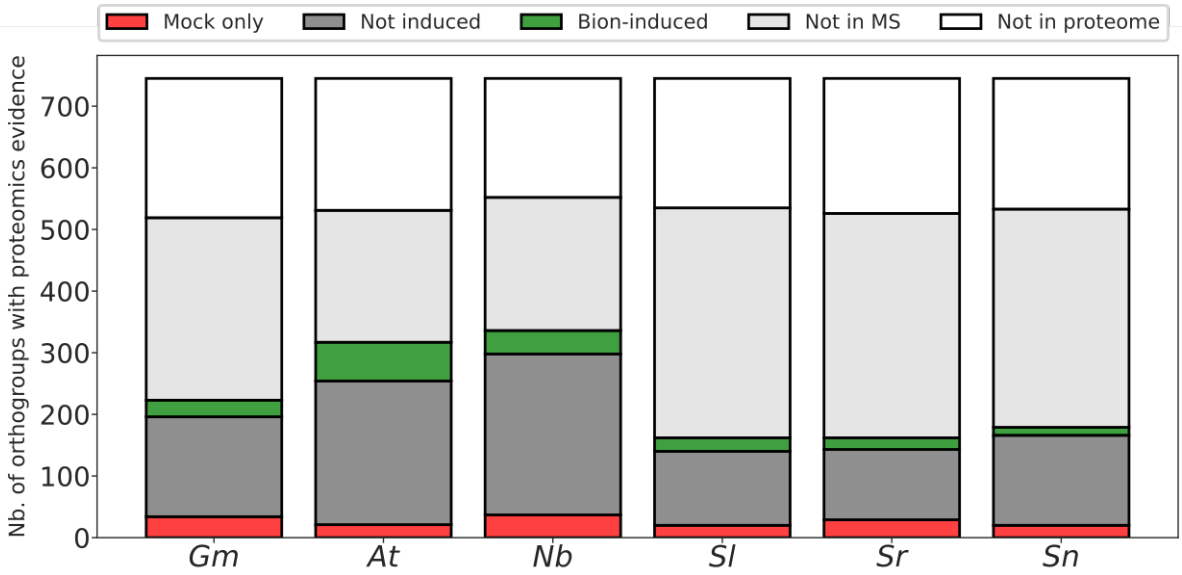


Figure 29 Apoplastic proteomes differ in size between species. All 745 orthogroups with proteomics evidence found between *G.max* (*Gm*), *A.thaliana* (*At*), *N.benthamiana* (*Nb*), *S.lycopersicum* (*Sl*), *S.peruvianum* (*Sr*) and *S.pennellii* (*Sn*). orthogroups were divided into the five categories of orthogroups detected in mock-treated samples only (Mock only), orthogroups detected in samples of both treatments and not induced by Bion (Not induced), Bion-induced orthogroups (Bion-induced), orthogroups not detected in AF samples of a species (Not in MS) and orthogroups not present in a species' proteome (Not in Proteome).

PCA analyses (compare **Figure 30** with **Figure 27**). The two components explaining most of the differences between all samples separate *A.thaliana* and *N.benthamiana* from *G.max, S.lycopersicum, S.peruvianum* and *S.pennellii* and explain species-specific differences of *N.benthamiana* and *A.thaliana* respectively as the remaining four species are centred (**Figure 30B**). PC1 and 2 are likely explaining orthogroups uniquely identified in apoplastic fluid extracts of *N.benthamiana* and *A.thaliana* as both contain approximately 100 orthogroups with proteomics evidence more than the other four species. PC3 separates *Solanaceae* tomato species and *N.benthamiana* from *G.max*, but only explains 6.87% of differences. Overall, PCA in orthogroup-based comparative proteomics explains more species-specific than treatment-induced differences.

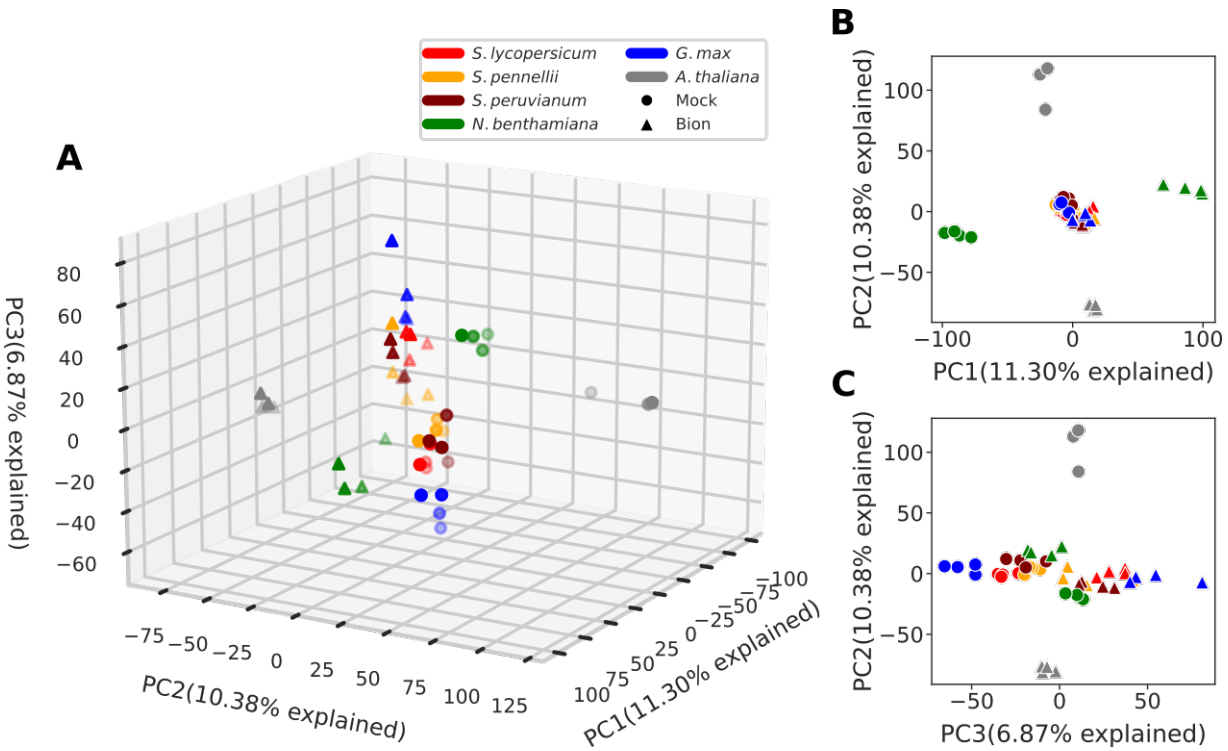


Figure 30 PCA of Bion-induced apoplastic proteomes mostly explains between-species differences. 3D (**A**) and 2D (**B, C**) principal component analysis of robustly detected, predicted to be secreted protein groups of *G.max, A.thaliana, N.benthamiana, S.lycopersicum, S.peruvianum* and *S.pennellii*.

3.3.3.2 Comparing PR protein accumulation between protein gels and MS data

To compare if MS data mirrored PR protein accumulation, we compared protein gels of Bion-induced AFs to orthogroups with the highest LFQ intensity for *A.thaliana, G.max* and *N.benthamiana*. For *N.benthamiana*, expected protein accumulation was based on *N.tabacum* and *N.sylvestris* which have highly similar patterns of PR protein accumulation (Parent et al. 1985). Both accumulate PR1 (14-17 kDa), PR3 (22-26 kDa) and PR2 (35-38 kDa) (Parent et al. 1985; Jamet and Fritig 1986; Legrand et al. 1987) and we observed Bion-induced protein accumulation of corresponding molecular weight on protein gels for *N.benthamiana* (**Figure 31A**). Band I matched a Bion-induced PR2 orthogroup with high intensity and band III matched a Bion-induced PR1 orthogroup with high intensity (**Figure 31A**). Band II, which is expected to have chitinase activity (Legrand et al. 1987), matched the molecular weight of Bion-induced metalloprotein PR17 (**Figure 31A**).

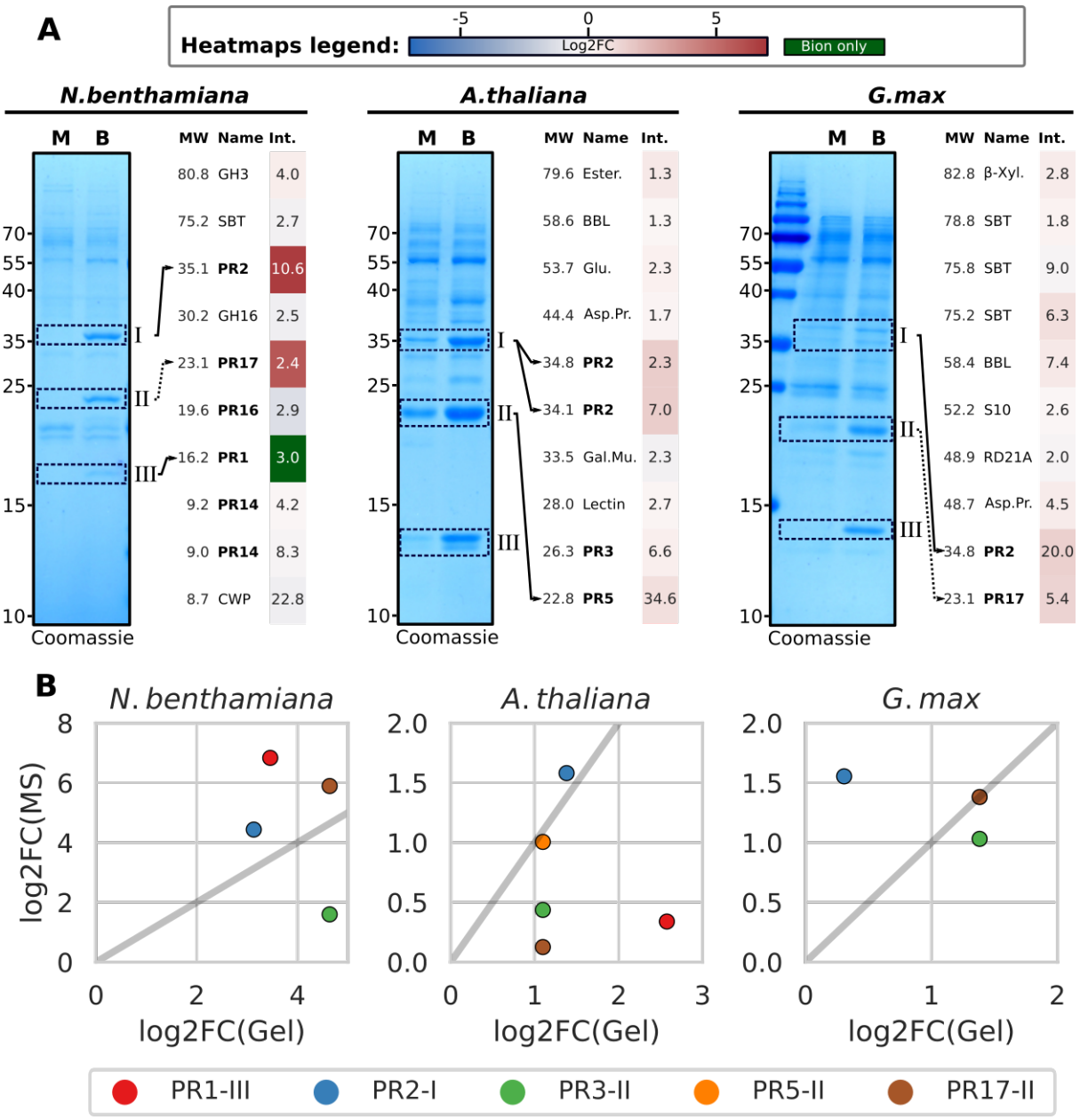


Figure 31 Orthogroup-based mass spectrometry data describes PR protein accumulation. (**A**) Cropped coomassie-stained SDS-PAGE gels of mock- (M) and Bion-induced (B) AFs of soybean, *A.thaliana* and *N.benthamiana*. Dashed boxes labelled I, II and III indicate accumulating PR proteins which were quantified using ImageJ and correlated with log2FC of MS data in **B**. Values in heatmaps represent the mean percentage shares of intensity of an orthogroup in Bion-induced AFs and colours encode Bion-induced log2FCs. Molecular weight (MW) is in kDa and without signal peptides. (**B**) Correlation of Bion-induced log2FC of MS and gel band intensities. The grey lines indicate a perfect 1:1 correlation. β-Xyl: β-xylosidase, GH: glycosylhydrolase, Ester.: esterase, SBT: subtilase, BBL: berberine bridge-like, Glu.Glu.: glucoside glucohydrolase, Asp.Pr.: aspartic protease, Gal.Mu.: galactose mutarotase-like, CWP: cell wall-binding protein.

Recently published work, however, showed that *N.benthamiana* PR17 runs below band II (Morimoto et al. 2022). As for tomato species in **chapter 2**, we detected Bion-induced

25.0 and 27.0 kDa PR3 chitinases at low shares of total LFQ intensity (0.7% and 0.5% respectively). For *A.thaliana*, we expected apoplastic, biotic stress-induced accumulation of PR1 (14-17 kDa), PR5 (22-26 kDa) and PR2 (35-38 kDa) (Uknes et al. 1992) and we observed Bion-induced protein accumulation of corresponding molecular weights on protein gel (**Figure 31A**). Band I matched two Bion-induced PR2 orthogroups with high LFQ intensity (**Figure 31A**) and the two orthogroups encoded both known SA-induced *At*PR2s At3g57240 (NP_191283.2) and At3g57260 (NP_001325845.1/NP_191285.1) (Datta and Muthukrishnan 1999; Doxey et al. 2007). Band II matched high-LFQ intensity PR5, which doubled in LFQ intensity upon Bion treatment. This increase, however, was not statistically significant, likely due to high variation in mock-treated samples (multiple-testing corrected p-value of two-sided, unpaired t-test: 0.100). We detected PR1 at low share of LFQ intensity (1.2%) and its 27% increase was statistically not significant. For *G.max*, we could not identify extensive proteomic characterisation of biotic stress induced AFs as for *Nicotiana* species or *A.thaliana*. Based on expression data and whole leaf protein extracts, however, we expected increased abundance of PR1 (14-17 kDa), chitinases (22-26 kDa) and PR2 (35-38 kDa) (Graham et al. 2003) and we observed Bion-induced protein accumulation of corresponding molecular weights on protein gels (**Figure 31A**). Band I matched a Bion-induced and high-LFQ intensity PR2 orthogroup (**Figure 31A**). Band II matched the molecular weight of the high-LFQ intensity metalloprotein PR17 orthogroup, but no high-LFQ intensity chitinase orthogroups (**Figure 31A**). We did detect proteins from four different Bion-induced PR3 orthogroups (27.0-28.1 kDa), but at very low share of LFQ intensity (all <0.1%) and one low-LFQ intensity Bion-induced PR1 orthogroup (15.0 kDa) in *G.max* AFs. In conclusion, high shares of LFQ intensities of some PR proteins (PR2s and PR5) was indicative of their high accumulation on protein gels while low LFQ intensities of others (PR3s and PR1s) was not. To investigate if low PR1 and PR3 LFQ intensity was due to an insufficient number of distinct peptides detected in MS, we predicted their tryptic peptides and determined their identification by MS for *N.benthamiana, A.thaliana* and *G.max*. For tryptic peptide prediction with unipept v3.0.2 (Verschaffelt et al. 2021), we used full length protein sequences and a tryp-

tic peptide was considered identified if the full peptide, a semi-tryptic peptide subset or a peptide with a missed tryptic cleavage site containing the full predicted tryptic peptide were identified in MS. The identification workflow was standardised in the script 'peptides overview.sh' (Homma 2022d). In **Table 7**, numerators describe the number of predicted tryptic peptides identified in MS data and denominators describe the number of predicted tryptic peptides. For PR1s, we identified between 25.0% (2/8) and 71.4% (5/7) of all predicted tryptic peptides and we identified the lowest number of tryptic peptides (2/8) for proteins of the PR1 orthogroup with the highest share of LFQ intensity (*N.benthamiana*) (**Table 7**). For 25.0 to 28.1 kDa PR3s, we identified between 6.7% (1/15) and 80.0% (8/10) of all predicted tryptic peptides in MS data (**Table 7**). Overall, the number of identified tryptic peptides in MS data of *N.benthamiana* and *A.thaliana* was comparable to high-intensity PR2 in tomato in **chapter 2** (**Table 3**). We concluded that a lack of distinct tryptic peptides detected in MS did not provide an explanation for low shares of LFQ intensity for PR1s and PR3s. Biases in quantification of LFQ intensities could stem from different ionisation efficiencies or glycosylation of peptides (Müller and Winter 2017; Dincer et al. 2022).

Table 7: Fraction of predicted tryptic peptides identified in MS data

Orthogroup	MW [kDa]	*N.benthamiana*	*A.thaliana*	*G.max*
PR1	15.0	-	-	5/7
PR1	15.2	-	4/9	-
PR1	16.2	2/7, 2/8	-	-
PR3	25.0	8/10	-	-
PR3	26.3	-	9/15	-
PR3	27.0	6/14	2/13	-
PR3	27.5	-	6/13, 6/12, 6/11	1/15
PR3	27.6	-	-	2/12
PR3	28.1	-	-	1/15

MW: molecular weight

To investigate the correlation between induction of PR proteins on protein gels and MS data, we quantified protein accumulation from protein gels using ImageJ and correlated increasing signal intensities of protein bands with increasing LFQ intensities of PR1, PR2, PR3, PR5 and PR17 next. To calculate PR protein induction in MS data, we added up

non-transformed LFQ intensities of protein groups detected in orthogroups annotated as respective PR proteins and then calculated their log2 fold changes (log2FCs). Band I correlated well with Bion-induced increases in PR2 orthogroup intensity for *N.benthamiana* and *A.thaliana* and was higher in MS- compared to gel intensity data for *G.max* (**Figure 31B**). Changes in gel intensity of band II matched Bion-induced increase in LFQ intensity for PR5 in *A.thaliana* perfectly (**Figure 31B**). For band II in *N.benthamiana*, increase in LFQ intensity of PR3s was lower compared to increase in gel intensities and similar for *G.max* (**Figure 31B**). For *G.max*, correlation between increase in MS- and gel intensities between band II and PR17 was high. PR1 increased stronger in abundance in MS data for *N.benthamiana* compared to band III gel intensity and less for *A.thaliana* (**Figure 31B**). For *G.max*, log2FCs for PR1 could not be calculated since it was only found in Bion-induced samples which supports strong Bion-induction. Taken together, increases of LFQ intensity in MS data did not directly reflect increases of matching protein bands for *N.benthamiana, A.thaliana* and *G.max* well. Differences could stem from other proteins of similar size overlapping with studied PR proteins in either Mock- or Bion-treated samples on protein gels.

Comparison to the tomato MS data (**chapter 2**) showed several common patterns and tomato-specific differences in Bion-induced apoplastic fluid extracts. Common was PR2 protein accumulation at 35-38 kDa and different orthogroups had the highest share of intensity in *Solanaceae* species compared to *A.thaliana* and *G.max* (**Figure 32**). Furthermore, Bion-induced PR1 induction was common across all six species and we observed distinct PR1 orthogroup accumulation in tomato species, *G.max* and *A.thaliana* (**Figure 32**). Moreover, PR3 accumulated in Bion-induced AFs across all species except for *A.thaliana* and PR3 chitinases displayed diverse induction of different orthogroups across species (**Figure 32**). PR9 peroxidases accumulated in Bion-induced AFs across species and showed highest shares of LFQ intensity in *N.benthamiana*, *S.lycopersicum* and *S.pennellii* (**Figure 32**). A clear distinction between *Solanaceae* and *G.max* and *A.thaliana* was Bion-induced accumulation of PR7 subtilases which showed highest shares of LFQ intensity in tomato species (**Figure 32**). Last, PR5 as the bona-fide marker of

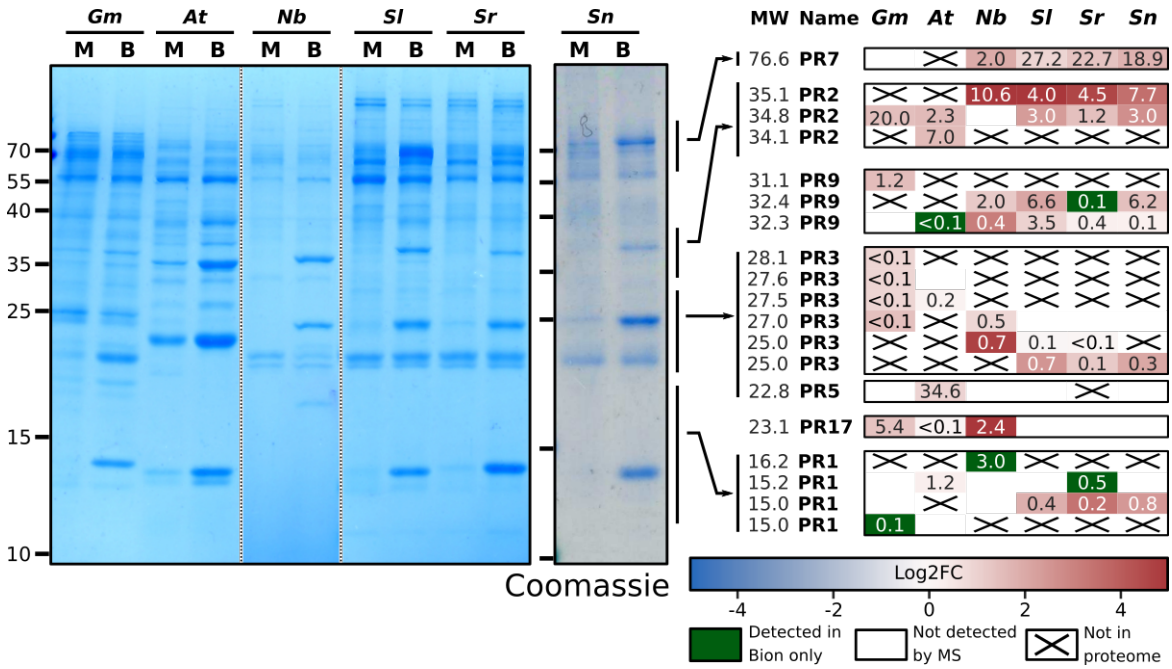


Figure 32 Orthogroup-based MS data reflects protein gels of six plant species. Cropped coomassie-stained SDS-PAGE gels of mock- (M) and Bion-treated (B) apoplastic fluid extracts of soybean (*Gm*), *A.thaliana* (*At*), *N.benthamiana* (*Nb*), *S.lycopersicum* (*Sl*), *S.peruvianum* (*Sr*) and *S.pennellii* (*Sn*). All samples were run on the same gel except for *S.pennellii*. Bars next to the gel of *S.pennellii* indicate the position of the corresponding molecular weight ruler on the left. Values in the heatmap represent the mean percentage share of LFQ intensity of an orthogroup in Bion-induced AFs and colours encode Bion-induced log2FCs. Molecular weight (MW) is in kDa and without signal peptides.

SAR was only detected via mass spectrometry in *A.thaliana* (van Loon et al. 2006).

3.3.3.3 Comparison of abundant functional groups in Bion-induced secretomes

To analyse prominent differences between Bion-induced AFs, we used LFQ-normalised values as a proxy for protein abundance and each orthogroup's share of total LFQ intensity as a proxy for its relative protein concentration. This was based on analyses from the original MaxLFQ paper concluding that MaxLFQ-derived LFQ values are a proxy for absolute protein abundance (Cox et al. 2014). In line with their conclusion, our comparison of gel- and LFQ intensity highlighted that highly abundant proteins such as PR2s, PR5 and PR7 were detected at high LFQ intensities. We furthermore discovered consistent biases in MS quantification for PR1s and PR3s and thus expect to consistently underestimate the abundance of proteins with biased MS detection following this approach.

In all six species, we saw that a few orthogroups defined the majority of LFQ intensities in Bion-induced apoplastic fluid extracts (**Figure 33**). Only 8-18% of orthogroups were

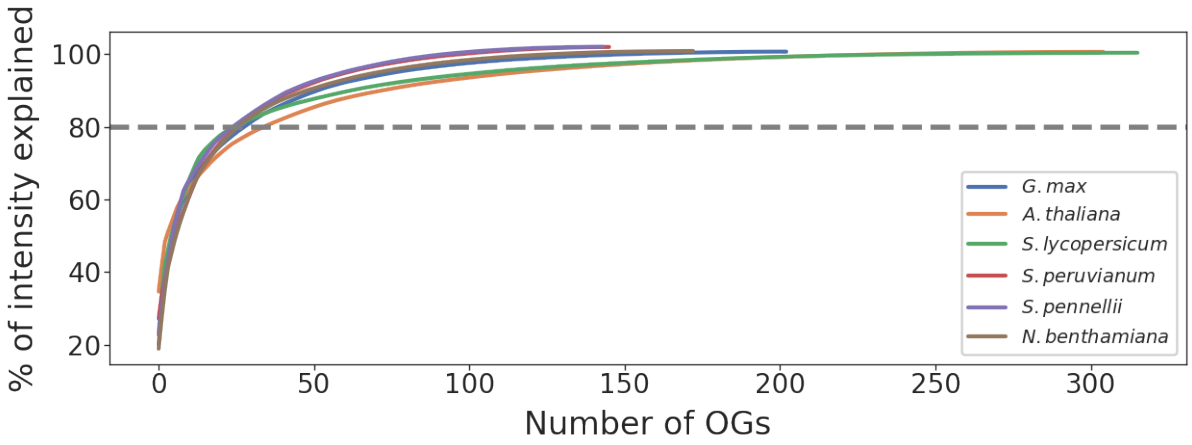


Figure 33: Few orthogroups define the majority of LFQ intensity in apoplastic fluid extracts. The dashed horizontal line indicates 80% of the total LFQ intensity of a species explained by the fewest number of orthogroups possible.

necessary to explain 80% of total LFQ intensity in Bion-induced AF extracts. Since less then 20% of orthogroups explained 80% of the protein LFQ intensity, we called these orthogroups pareto orthogroups, based on the pareto principle which states that approximately 80% of outcomes can often be explained by 20% of causes. Each species had its own set of pareto orthogroups which were similar in size for *G.max* (29), *N.benthamiana* (25), *S.lycopersicum* (26) *S.peruvianum* (25) and *S.pennellii* (27) and larger for *A.thaliana* (35). Together, pareto orthogroups consisted of 106 of all 745 apoplastic orthogroups with proteomics evidence. To understand the composition of Bion-induced AFs defined by pareto orthogroups, we manually added functional annotations to all 106 orthogroups, using NCBI annotations supplemented with Pfam annotations. Comparing shares of total LFQ intensities as proxies for concentration of orthogroups defining most of the apoplastic proteomes of the six species illustrated distinct differences between species and highlighted similarities between tomato species (**Figure 34**). All species showed glycosyl hydrolases and ROS-related orthogroups such as PR9 peroxidases and PR16 oxalate oxidase-like proteins at high relative abundance. Another strong difference was variation of relative abundance of proteases, especially S8 subtilases between plant species. In the

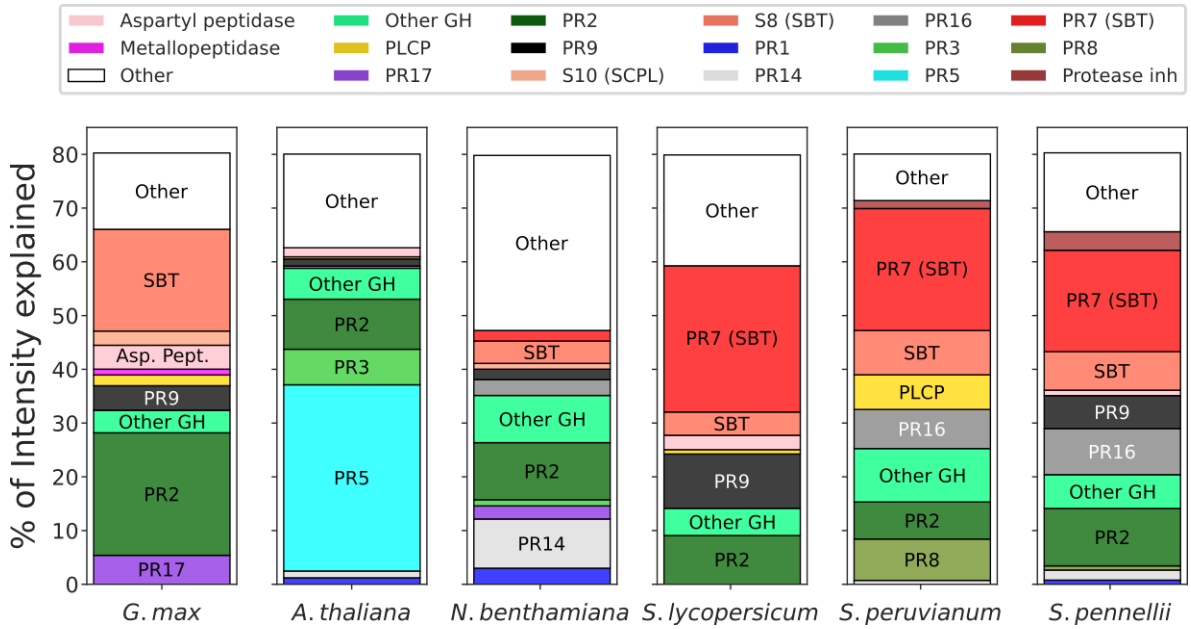


Figure 34: Proteomes of apoplastic fluid extracts differ between species. Clusters of functional annotations in pareto orthogroups and their mean share of total samples LFQ intensities. Clusters >4% were annotated in the bar chart.

next section, we will discuss ROS-related PR9 and PR16 orthogroups as well as proteases.

3.3.3.4 Peroxidases and proteases as further prominent pareto orthogroup clusters in Bion-induced AFs

First, we analysed reactive oxygen species (ROS)-associated PR9 peroxidases and PR16 oxalate oxidase-like/germin-like proteins which either generate or sequester ROS (Lane 1994; Shigeto and Tsutsumi 2016; Pei et al. 2019). While shares of LFQ intensity of PR16 was high in Bion-induced AFs, its accumulation was not induced by Bion treatment in any of the six species. On the contrary, shares of LFQ intensities approximately halved in *S.peruvianum, S.pennellii* and *N.benthamiana* upon Bion treatment which was in stark contrast to PR9 peroxidases. Apoplastic PR9 peroxidases were Bion-induced and one *Solanaceae*-specific orthogroup contributed most relative PR9 LFQ intensity in *S.lycopersicum* (6.6%), *S.pennellii* (6.2%) and *N.benthamiana* (2.0%), but was mostly absent from *S.peruvianum* (0.2%) (**Figure 35**). This PR9 orthogroup is encoded by a single protein in each species (Solyc01g006300.3.1, SperuGCA_000102, XP_015065492.1 and NbD024399.1). Furthermore, another Bion-induced PR9 orthogroup encoding pro-

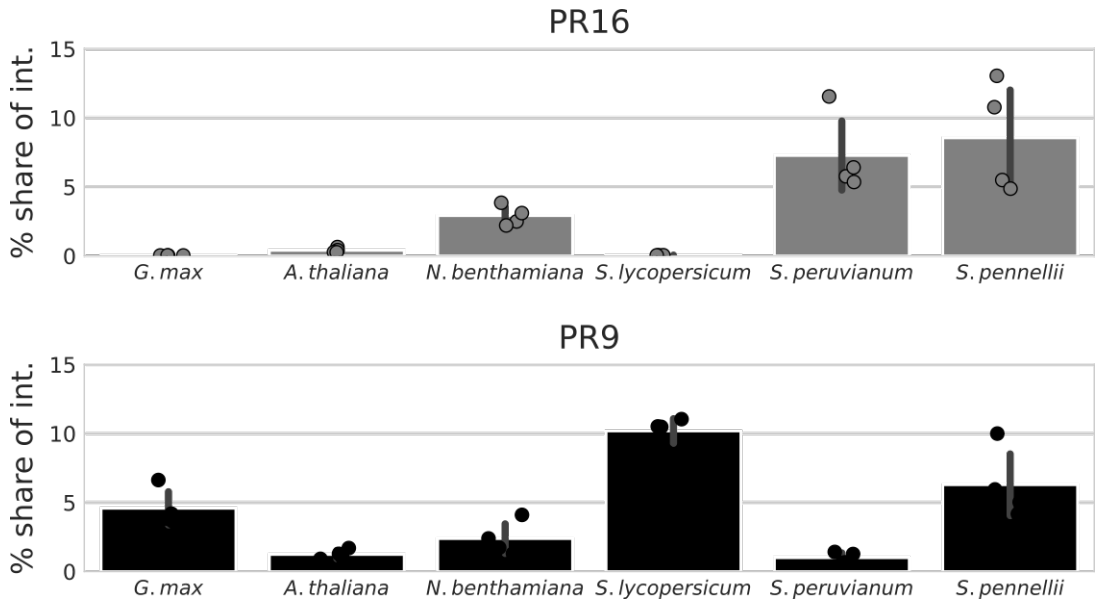


Figure 35: ROS-associated proteins in Bion-induced apoplastic fluids. Shares of total sample LFQ intensities in AFs of Bion-treated plants. Bar charts display the mean of share of LFQ intensity in Bion-induced AF and error bars indicate standard deviation.

teins in all six species was predominantly detected in *S.lycopersicum* (3.5%). This PR9 (Solyc01g105070.3.1) is under selection in tomato species (**Figure 22** in **chapter 2**). Interestingly, the two PR9 pareto orthogroups most abundant in Bion-induced AFs of *G.max* were species-specific, indicating that *G.max* secretes a unique profile of peroxidases compared to the other five species.

Next, we compared apoplastic proteases encoded by pareto orthogroups of different Bion samples (**Figure 36**). These included C01 Papain-like cysteine proteases (PLCPs), A01 aspartic proteases, a M10 matrix metalloprotease and S10 serine carboxypeptidase-like (SCPL) proteases (Rawlings et al. 2018). S8 subtilases will be discussed in detail later. In tomato species, two pareto orthogroups encoded PLCPs in *S.lycopersicum* and *S.peruvianum* and PLCPs made up 6.4% of total sample LFQ intensities in *S.peruvianum* (**Figure 36A**). The homolog of Pip1 (SperuGCA_000063) was the most prominent (5.3%) and a RD19A homolog (SperuGCA_057723) was the second most prominent (1.1%) PLCP in *S.peruvianum*. The same RD19A orthogroup had the highest share of LFQ intensity of PLCPs in Bion samples in *S.lycopersicum* (Solyc01g110110.3.1) and *S.pennellii* (XP_015064040.1) at 0.8% and 0.6%, respectively (**Figure 36A**). Two different PLCP or-

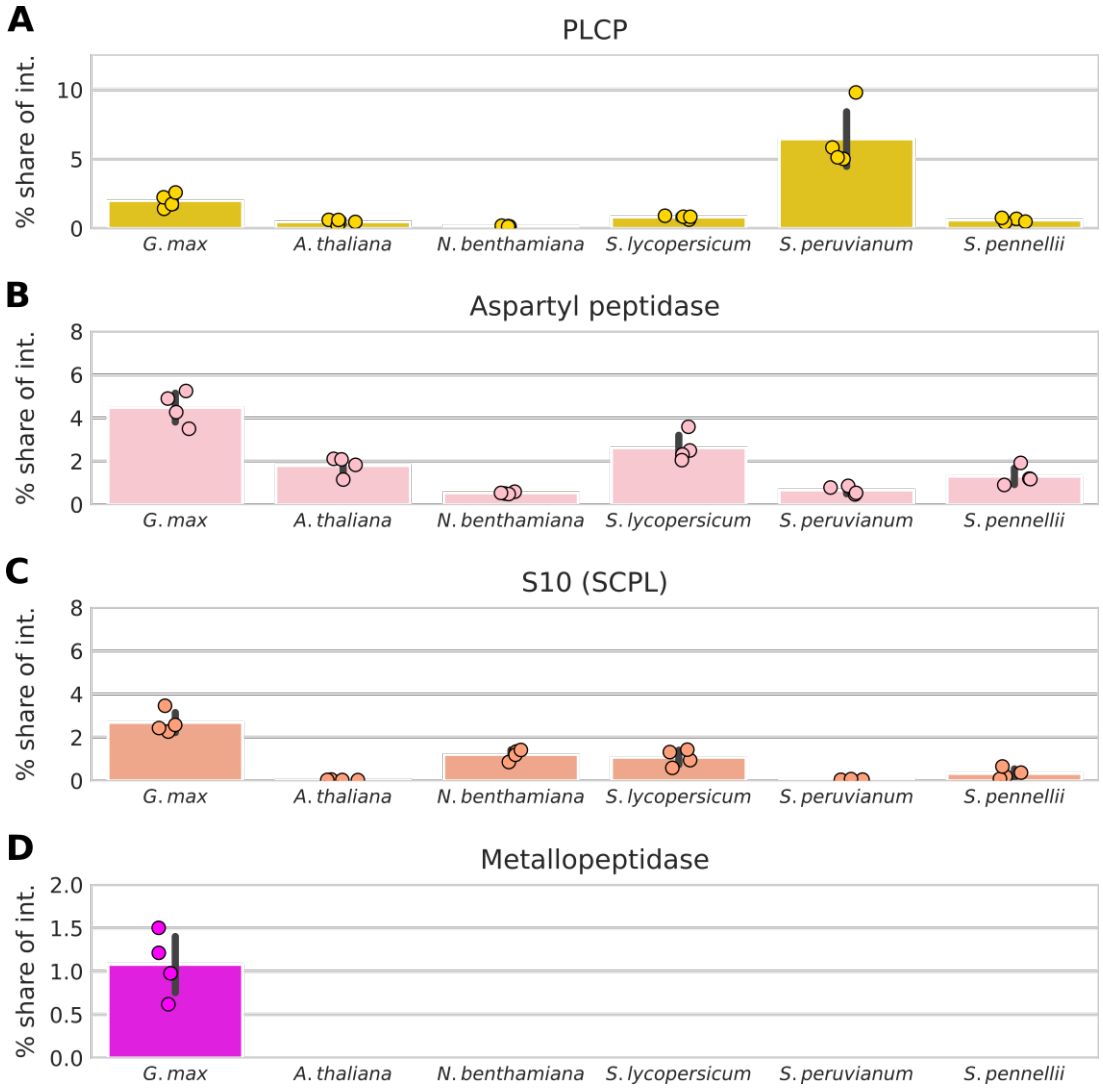


Figure 36: Similarities and differences between protease landscapes of different species. Shares of total sample LFQ intensities in AFs of Bion-treated plants. Bar charts display the mean of share of LFQ intensity in Bion-induced AF and error bars indicate standard deviation. PLCP: Papain-like cysteine protease, SCPL: Serine-carboxypeptidase-like protease.

thogroups encoding RD21A and aleurain-like homologs were most prominent in *G.max* (2.0%) and *A.thaliana* (0.4%), respectively (**Figure 36B**). Furthermore, we detected aspartic proteases in pareto orthogroups of Bion samples with high shares of total LFQ intensities for *G.max* and *S.lycopersicum* (**Figure 36B**). In *S.lycopersicum*, two pareto groups with each one aspartic protease were detected at 1.1% (Solyc08g068870.3.1) and 1.5% (Solyc04g076190.1.1) share of LFQ intensity each. The latter orthogroup was also detected in Bion-induced AFs of *S.pennellii* (1%), *S.peruvianum* (0.4%) and *A.thaliana* (1.7%). In *G.max*, a single pareto orthogroup encoded three different aspartic proteases

(XP_003532146.1, XP_003532147.1 and XP_003551807.1), none of which encoded the immune protease GmAP1 (Guo et al. 2019). GmAP1 (XP_003546159.1) was only detected in mock samples at 0.2% share of LFQ intensity. Moreover, we detected different SCPL proteases in pareto orthogroups of *G.max* and *N.benthamiana* (**Figure 36C**). In *N.benthamiana* Bion-induced AF, we detected one pareto orthogroup encoding one SCPL (NbD026891.1) at 1.1% share of total LFQ intensity and another pareto orthogroup in *G.max* with two SCPLs (XP_006599137.1 and XP_006599138.1) at a combined share of LFQ intensity of 2.6%. Both orthogroups were also detected in *S.lycopersicum* at 0.5% and 0.6% share of LFQ intensity, respectively. Last, a single pareto orthogroup with one metalloprotease (NP_001235535.2) was detected in *G.max* (**Figure 36D**). Taken together, apoplastic PLCPs, aspartic proteases, SCPLs and metalloproteases were detected amongst the most abundant pareto orthogroups across Bion samples of the six different species. While few immuno proteases such as Pip1 were defining proteases of apoplastic landscapes, none were as abundant as S8 subtilases.

3.3.3.5 S8 Subtilases

Next, we compared functional clusters of S8 subtilases, specifically highlighting the P69 subtilase orthogroup (PR7). Within plant subtilases, the large P69 cluster has been described to be *Solanaceae*-specific (Taylor and Qiu 2017) which was mirrored in our data (**Figure 37**). Furthermore, high levels of P69s was a distinct feature of Bion-treated tomato AF extracts, as in *N.benthamiana*, the P69 subtilase orthogroup made up a 13.67 times lower share of LFQ intensity of apoplastic proteins compared to *S.lycopersicum* (**Figure 37**). In *N.benthamiana*, other S8 subtilases were more prominent than P69 subtilases, combining for 4.2% share of LFQ intensity (**Figure 37**). In *G.max*, no P69 subtilases were detected, but other S8 subtilases made up 18.9% share of LFQ intensity (**Figure 37**). Last, we only detected few subtilases in Bion samples of *A.thaliana* and their share of total LFQ intensities were below 1%. To analyse the seven different S8 pareto orthogroups in more detail, we returned to the complexity of individual S8 subtilases not seen with our orthogroup-based approach. Therefore, we analysed shares of total LFQ

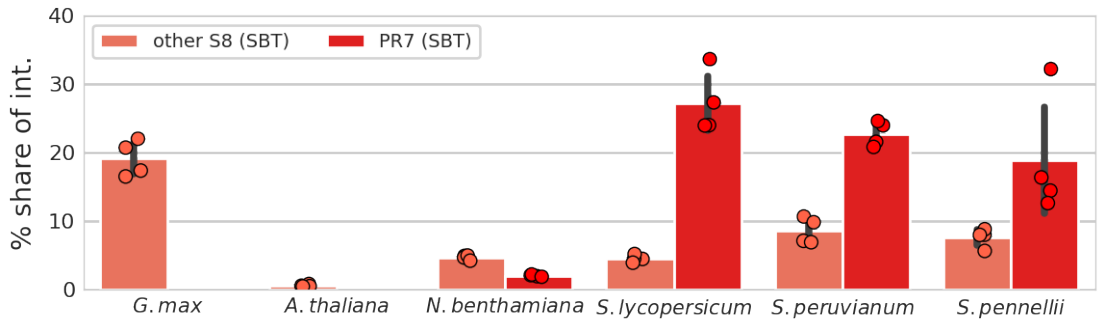


Figure 37 Different species have different apoplastic subtilase landscapes in Bion-induced immuno secretomes. Values represent shares of total sample intensities of subtilase orthogroups in four biological replicates per species. The subtilase orthogroup containing all tomato P69s was separately highlighted as PR7 subtilases. Bar charts display the mean of Bion-induced MS samples and error bars indicate standard deviation.

intensities of all proteins included in the seven S8 subtilase pareto orthogroups. Within these seven S8 pareto orthogroups, we detected 70 subtilases across the six species. To understand how the S8 proteases relate to each other, we generated a maximum-likelihood phylogenetic tree using whole CDSs for each protease and highlighted subtilase clades based on published annotations (Taylor and Qiu 2017; Reichardt et al. 2018). In the phylogenetic tree, we detected SBT4 and SBT1 clades (**Figure 38**). The SBT1 clade was further divided into a large SBT1.1-1.8 clade, a phytaspase-containing SBT1.13 clade and a P69 (PR7)-containing SBT1.10 clade. In Bion samples, SBT1.10s contained a large P69 clade and a smaller but distinct clade including *Sl*SBT1.10q (Reichardt et al. 2018). Based on shares of LFQ intensity, the *Solanaceae*-specific SBT1.10s contained the most abundant proteins of all subtilases and showed dynamic changes in abundances with three proteins decreasing in abundance and nine proteins increasing in abundance. P69B increased in abundance across all three tomato species and was the most abundant subtilase in *S.peruvianum* (SperuGCA_052100) and *S.pennellii* (XP_015085422.1) and the second most abundant subtilase in *S.lycopersicum* (Solyc08g079870.3.1, P69B) (**Figure 38**). The most abundant subtilase in Bion-induced *S.lycopersicum* AFs was P69A (**Figure 38**). P69G was only detected in Bion-induced AFs of *S.peruvianum* (SperuGCA_088264) and *S.lycopersicum* (Solyc08g079900.3.1) and increased in both species

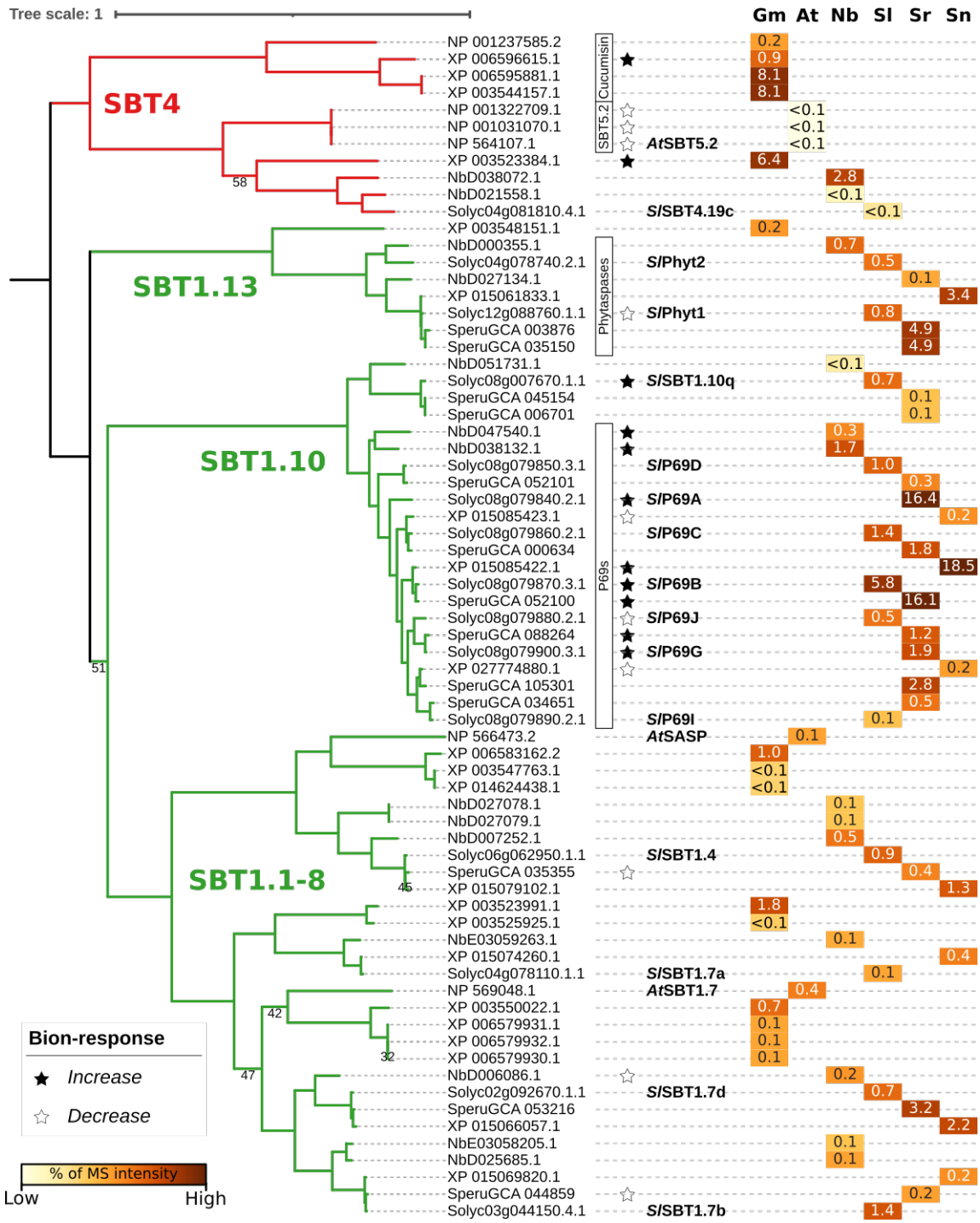


Figure 38 Maximum-likelihood phylogenetic tree of 70 subtilases found in pareto orthogroups. Functionally annotated subtilases and sub clades were highlighted and stars indicate increase or decrease in relative abundance of apoplastic proteins upon biotic stress trigger. A heatmap indicates the % share of LFQ intensity of each subtilase in soybean (*Gm*), *A.thaliana* (*At*), *N.benthamiana* (*Nb*), *S.lycopersicum* (*Sl*), *S.peruvianum* (*Sr*) and *S.pennellii* (*Sn*). MS protein groups containing multiple proteins (such as XP_006595881.1 and XP_003433157.1) were split for illustration purposes. The tree was based on aligned whole CDS nucleotide sequences of 70 subtilases from pareto orthogroups and only bootstrap values below 70 are displayed. Colouring of subtilase clades was based on Reichardt et al. 2018.

upon Bion treatment. Furthermore, two *N.benthamiana* subtilases in a distinct sister clade of P69s increased in abundance upon Bion treatment, one of which (NbD038132.1)

was the second most abundant subtilase in Bion samples of *N.benthamiana* (**Figure 38**). Its total abundance measured by its share of LFQ intensity at 1.7% was tremendously lower compared to highly-abundant P69s in tomatoes. Changes in abundance of P69A, P69B, P69C and P69G were in line with published MS data of apoplastic fluid extracts of *Ralstonia solanacearum*-infected *S.lycopersicum* (Planas-Marquès et al. 2018), confirming robustness of increasing subtilase abundance upon different biotic stress triggers across tomato cultivars and species. Next, we analysed the SBT1.13 clade, which contained two *S.lycopersicum* phytaspases, *Sl*Phyt1 and 2 (**Figure 38**) (Reichardt et al. 2018). Both phytaspases have distinct substrate specificity and transient over expression of *Sl*Phyt2 in *S.lycopersicum* triggers cell death (Reichardt et al. 2018) and Phyt2 was only detected in Bion-induced AFs of *S.lycopersicum*, but not in AFs of other tomato species. Furthermore, Phyt1 decreased in abundance in *S.lycopersicum* while remaining unchanged in *S.peruvianum* and *S.pennellii* (**Figure 38**). Wild tomato *Sl*Phyt1 orthologs had nearly six- and more than four-fold relative share of LFQ intensity in *S.peruvianum* and *S.pennellii* compared to *S.lycopersicum*, indicating differences in apoplastic phytaspase composition between wild and domesticated tomato species. Furthermore, the large and diverse SBT1.1-SBT1.8 cluster contained many uncharacterised subtilases from all six species (**Figure 38**). Abundances of subtilases in this cluster either remained unchanged or decreased upon biotic stress signalling. Next, we analysed the SBT4 clade which included the most abundant *N.benthamiana* subtilase (NbD038072.1) at 2.8% share of Bion sample LFQ intensities. SBT4 also contained the two most abundant *G.max* subtilases, one characterised as cucumisin (NCBI annotation) and the other being uncharacterised (**Figure 38**). The SBT4 clade also included three isoforms of *At*SBT5.2 (**Figure 38**). *At*SBT5.2 had been placed in the SBT4 clade in larger-scale maximum-likelihood phylogenetic analyses of plant subtilases before (Taylor and Qiu 2017). Transcripts of apoplastic *At*SBT5.2 were shown to be down-regulated upon bacterial infection (Serrano et al. 2016), consistent with *At*SBT5.2s decreasing in abundance upon Bion treatment (**Figure 38**). We did not detect apoplastic immuno hydrolases SBT3 from *S.lycopersicum* or SBT3.3 from *A.thaliana* in Bion samples (Ramírez et al. 2013; Meyer et al. 2016). Both immuno

subtilases are responsive to pathogen-infection, *Sl*SBT3 additionally being responsive to wounding and *At*SBT3.3 being explicitly salicylic-acid independent (Ramírez et al. 2013; Meyer et al. 2016). Our data thus stressed trigger-specificity of different subtilase clades. Taken together, a detailed analysis of subtilases in Bion-induced AFs across six plant species highlighted three key aspects: First, P69 subtilases were only found in *Solanaceae* species and dominated tomato apoplasts where they accumulated upon Bion treatment. Second, Bion samples of distantly-related wild tomato species contain higher shares of cell death-related phytaspases than domesticated *S.lycopersicum*. Third, the Bion samples of *G.max* contained high shares of yet-poorly characterised SBT4 subtilases. Subtilases marked stark differences in apoplastic immunity landscapes which pathogens have to navigate to successfully infect different plant species.

3.4 Discussion

In this chapter, we explored similarities and differences between biotic stress-induced AFs of six plant species. We expanded the repertoire of Bion-induced apoplastic proteomes from the previous tomato chapter with three eudicot species *G.max, A.thaliana* and *N.benthamiana*. Comparison between different Bion-induced secretomes identified key similarities and differences in apoplastic immunity landscapes that pathogens have to navigate when attempting to colonise different plant species.

3.4.1 Preparing MS data for comparative proteomics

PCA of MS data of AFs had highlighted that one mock-treated sample of *A.thaliana* introduced more within-treatment difference than observed for the other five species. This sample was removed from further comparison as it lacked blocks of lower-abundance protein groups shared between all other *A.thaliana* samples. Notably, this was the same sample with increased MDH activity indicating a higher cytoplasmic contamination which likely translated into abundant cytoplasmic proteins masking the detection of lower-abundant apoplastic proteins. In future, AFs with multiple-fold increased MDH activity compared to other samples should be treated as contaminated and deemed not suitable for further characterisation.

3.4.2 Comparing PR protein accumulation between protein gels and MS data

LFQ intensities of orthogroups containing PR2s, 5 and 7 were consistently amongst the ten orthogroups with highest LFQ intensities per species. These PR proteins also consistently accumulated visibly to high degrees on protein gels. For PR1s and PR3s however, we identified mismatches between high gel- and low LFQ intensities, stressing biases in LFQ intensity-based quantification of proteins. Disproportionately low PR3 LFQ intensity could stem from glycosylated peptides. In *N.tabacum* and *N.sylvestris*, PR protein bands of size comparable to PR3s were retained by concanavalin A columns indicative of glycosylation (Parent et al. 1985). Concanavalin A is a jack bean-derived lectin binding specifically to glycosyl and mannosyl residues of glycoproteins (Goldstein et al. 1965).

Furthermore, PR3s isolated from *N.tabacum* migrate on native gel electrophoresis as proteins of substantially higher molecular weight (Jamet and Fritig 1986), which is common for glycosylated proteins (Wang et al. 2020a). A gel-based stain to detect glycoproteins (Konat et al. 1984), however, did not identify purified *N.tabacum* PR proteins to be glycosylated (Jamet and Fritig 1986). Glycosylation of PR3s is suggested by multiple independent evidence and could explain PR3s low LFQ intensity.

For *G.max*, no analysis of apoplastic PR protein accumulation has been published yet. In our MS data, high-LFQ intensity PR17 correlated with a Bion-induced PR protein band and future studies could determine if band II is PR17 metalloprotein or PR3 chitinases. To determine the protein encoding this band, metalloprotein labelling of PR17 (Morimoto et al. 2022) could identify if PR17 migrates at comparable size to the PR protein band. If both bands overlap, this PR protein band could be purified similar to PR proteins in *Nicotiana* species before (Jamet and Fritig 1986; Legrand et al. 1987) and tested for chitinase and metalloprotein activity.

3.4.3 Pareto orthogroups and their functional annotation

To understand key similarities and differences between Bion-induced AFs, we compared functional annotations of orthogroups defining the majority (80%) of LFQ intensities. A similarity between all six species was high LFQ intensity of PR2 β-1,3-glucanases which provide resistance against a broad range of fungi and oomycete (van Loon et al. 2006; Doxey et al. 2007; Balasubramanian et al. 2012) indirectly by callose deposition and directly by degrading β-1,3-glucans in fungal and oomycete cell walls, reducing structural integrity and releasing immunogenic elicitors (Bowman and Free 2006; Balasubramanian et al. 2012; Fesel and Zuccaro 2016). Interestingly, one PR2 orthogroup accumulated in Bion-induced AFs across species except for *N.benthamiana*. The proteome of *N.benthamiana*, however, encodes a homolog (NbD012752.1) of this orthogroup, identifying it as a candidate to be implicated in plant immunity. Future transient over expression in combination with infection of pathogens such as *P.infestans* could identify if this *N.benthamiana* PR2 enhances resistance against oomycetes. Other glycosyl hydro-

lases such as chitinases, increase disease resistance against fungi by hydrolysing fungal chitin to reduce pathogen cell wall integrity and release immunogenic elicitors (Schlumbaum et al. 1986; Takashima et al. 2020). Several chitinases increased in abundance upon Bion treatment including one putative *N.benthamiana* chitinase from a PR3 orthogroup (NbD015584.1, 0.6% share of intensity) which has antibacterial properties *in vivo* and does not hydrolyse chitin (Godson. et al, *in preparation*). We detected one homolog of this antibacterial PR3 in AFs of *S.lycopersicum* (Solyc02g082930.3.1), identifying it as a putative antibacterial protein in the tomato apoplast. Antimicrobial activity of putative chitinases against bacteria are further supported by Bion-induced PR8 in *S.peruvianum* as PR8s (class III chitinases) have dual function as chitinases and lysozymes (Terwisscha van Scheltinga et al. 1995). Antibacterial PR3 from *N.benthamiana*, however, does not show lysosomal activity (Godson. et al, *in preparation*) and thus likely confers increased antibacterial resistance via a different mechanism. In addition to glycosyl hydrolases, we detected ROS-associated PR9 and PR16 orthogroups in Bion samples. One single-gene, *Solanaceae*-specific and Bion-induced PR9 orthogroup encoded the most abundant peroxidases in *S.lycopersicum, S.pennellii* and *N.benthamiana* and it was functionally annotated as a lignin-forming anionic peroxidase (NCBI annotation). Another Bion-induced PR9 accumulated in AFs of *S.lycopersicum* (Solyc01g105070.3.1, 3.5%) and is a homolog of NP_201217.1/AT5G64120/AtPRX71 which is involved in increased lignification as well (Shigeto et al. 2015). Lignin deposition in the extracellular space can form a barrier reducing bacterial motility and fungal penetration, both to improve plant disease resistance (Bhuiyan et al. 2009; Lee et al. 2019). While PR9 peroxidases have been linked to defence-related ROS bursts (Shigeto and Tsutsumi 2016), these Bion-induced peroxidases more likely scavenge ROS than contribute to its accumulation. Furthermore, we found clear differences in protease landscapes between different species. The immunity protease Pip1 was the most prominent PLCP in wild tomato *S.peruvianum* and yet less-characterised aspartic proteases were consistently amongst the most abundant proteases in different Bion samples. Moreover, detailed analyses of subtilases revealed that SBT4 subtilases accumulated in Bion-induced AFs of *G.max* in contrast to the highly-abundant SBT1.10

subtilases in tomatoes.

3.4.4 An emerging theme for apoplastic immunity proteins

Abundance and LFQ intensity of apoplastic proteins in Bion samples are strong indicators for involvement in plant immunity as shown by PR5 Thaumatin-like protein (Liu et al. 2012; Acharya et al. 2013; Misra et al. 2016), P69 subtilases (Paulus et al. 2020; Wang et al. 2021a), PR2 β-1,3-endoglucanases (Stintzi et al. 1993) and the PLCP Pip1 (Ilyas et al. 2015). Extrapolating from this observation, we hypothesise that high-LFQ intensity *G.max* SBT4 subtilases, tomato lignifying peroxidases and tomato PR8 chitinases are involved in apoplastic plant immunity. High abundance and LFQ intensity are however no necessary attribute of defence-related proteins. Rcr3 is 10- to 20-fold less abundant than Pip1 in BTH-induced and PLCP-enriched tomato AF (Ilyas et al. 2015) and BGal1 of *N.benthamiana* and BGal8 of *A.thaliana*, both involved in immunity against *P.syringae* (Buscaill et al. 2019; Buscaill et al. 2021), were detected at low LFQ intensities (0.04% and 0.05%) in our Bion-induced AFs. Regardless of abundance of defence-related proteins, pathogens need to overcome deployed immunity proteins to colonise a host species. This can be achieved by direct inhibition of hydrolytic activity and for immunity-related hydrolases listed above, pathogen-derived inhibitors such as glucanase inhibiting proteins for PR2 (Ham et al. 1997; Rose et al. 2002; Damasceno et al. 2008), Avr2 and EpiC2B for Pip1 (Rooney et al. 2005; Tian et al. 2007) and Epi1 and Epi10 for P69s (Tian et al. 2004; Tian et al. 2005) are known. It is noticeable that these inhibitors target highly-abundant proteins and were first detected from pathogens colonising hosts in which their targets were relatively more abundant than in other species as seen in our LFQ intensity data set for PR2 in *G.max* and Pip1 and P69s in tomatoes. Following this logic, we predict to find SBT4 subtilase and aspartic protease inhibitors in pathogens infecting *G.max* and we expect PR8 class III chitinase inhibitors produced by tomato pathogens. While inhibition of low-LFQ intensity immune hydrolases such as BGal1 from *N.benthamiana* (Chandrasekar et al. 2017) or Rcr3 from tomato (Rooney et al. 2005; Song et al. 2009; Shindo et al. 2016) are described, no proteinaceous inhibitor of chitinases has been dis-

covered yet. Various other mechanisms to suppress chitinase-associated plant immunity include pathogen modification of chitin by deacetylation to generate the less immunogenic chitosan (Gao et al. 2019a), hiding and sequestering of chitin through binding of Avr4 (van den Burg et al. 2006) and Ecp6 (de Jonge et al. 2010), degrading released chitin by pathogen-released chitinases (Martínez-Cruz et al. 2021) and directly cleaving plant-derived chitinases (Naumann et al. 2011; Jashni et al. 2015; Han et al. 2019) (recently reviewed in (Buscaill and van der Hoorn 2021)). Our data highlighted multiple different Bion-induced chitinases, which, in combination with the diverse mechanisms to evade chitinase-mediated immunity, culminates in a tentative hypothesis: diverse and potentially redundant defence mechanisms such as chitinases might evolutionarily be overcome robustly by indirect measures such as hiding elicitors from the host while less diversified defence mechanisms are overcome by direct interference. Direct interference could encompass direct inhibition by a pathogen-derived inhibitor, but also specific suppression of secretion into the apoplast as described for the *Phytophthora* effectors *Pi*Avrblb2 and *PsAvh240* which suppress the secretion of immune proteases (Bozkurt et al. 2011; Guo et al. 2019). Identification of a plant-derived chitinase inhibitor would weaken this hypothesis.

3.4.5 Apoplastic defence proteomes induced by different elicitors

In this chapter, we studied biotic stress-induced AFs, by stimulating the SA signalling pathway which is central to systemic acquired resistance against biotrophic pathogens (Gao et al. 2015). During an infection however, a plant is exposed to a range of different elicitors such as damage-/microbe-/pathogen-associated molecular patterns (D/M/-PAMPS) and pathogen effectors which trigger defence responses (Abdul Malik et al. 2020). A question thus emerges: how conserved are abundant, Bion-induced proteins across different biotrophic defence-induced secretomes within the same species? To study the full defence protein repertoire of a plant species, AFs induced by different elicitors such as flagellin, chitin, peptidoglycan or H_2O_2 treatment could be studied in a pan-proteomics approach. Furthermore, analysing AFs from pathogen-challenged plants can help identify

proteins increasingly secreted during actual pathogen infections. Data sets of biotic stress-induced AFs of *N.benthamiana* infected with *A.tumefaciens* (Grosse-Holz et al. 2018) and *S.lycopersicum* infected with *R.solanacearum* (Planas-Marquès et al. 2018) are already available. Furthermore, studying Bion-induced AFs focuses on defence responses against biotrophic pathogens, but some of the most important plant pathogens have necrotrophic stages in their life cycles including hemibiotrophic oomycetes *Phytophthora infestans, sojae* and *ramorum* or necrothopic fungi *Botrytis cinerea, Mycosphaerella graminicola* and *Colletotrichum* species (Dean et al. 2012; Kamoun et al. 2015). Effectors of necrotrophic pathogens target and hijack disease resistance pathways against biotrophic pathogens culminating in programmed cell death to promote colonisation of host tissue (Shao et al. 2021). This effector-triggered susceptibility turns the plant immune system into a plant susceptibility system. While induced systemic resistance against necrotrophic pathogens, mediated by jasmonic acid (JA) and ethylene, is not associated with systemic PR protein accumulation (van Loon et al. 2006; Romera et al. 2019), it would be interesting to compare systemic adaptations of apoplastic proteomes of JA-treated plants to apoplastic proteomes of mock- and Bion-induced plants. This comparison would highlight putative similarities and contrasts in systemic plant defence preparations against pathogens with fundamentally different life styles.

3.4.6 Outlook

In future, MS analyses of biotic stress-induced secretomes of more species could generate a map of apoplastic immunity landscapes. Studying closely-related species which can be infected by the same pathogens could define the range of apoplastic immunity landscapes a certain pathogen is capable of navigating. Further expanding to more distantly-related, non-host species can help to identify apoplastic proteins defining an environment a given pathogen can not survive anymore. These apoplastic proteins could then be studied for their implication in plant immunity by expressing them in current host plants or depleting them in non-host plants followed by pathogen infection assays. These experiments could help understand which proteins contribute to non host resistance of closely-related

species. Intuitively, *Solanum* species which are amenable to experimental procedures such as biotic stress induction and apoplastic fluid extraction, with large publicly available genomic resources and agriculturally important crops such as tomato (*Solanum lycopersicum*), potato (*Solanum tuberosum*) and aubergine (*Solanum melongena*) are ideal starting points. Then, mapping of apoplastic immunity landscapes could be extended to more distantly-related *Solanaceae* species to close the gap to *N.benthamiana* data generated for this chapter.

To compare apoplastic proteomic landscapes between species, an orthogroup-based approach has proven robust. There are limitations and complementary approaches to engineering apoplastic immunity landscapes, such as resistance gene breeding from closely-related species. Not for all crops, however, are highly diverse, qualitative resistance trait-rich wild relatives accessible for breeding. One such example is sugar beet (Monteiro et al. 2018). Engineering apoplastic environments by over expressing immunity-related proteins from other plant species could increase plant immunity in such crops. Caveats of transferring apoplastic immunity proteins between species need to be considered. Downstream signalling components might not be conserved across species and defence proteins might not be stable in apoplastic environments shaped by different proteases. For instance, apoplastic *Solanum* proteins are under evolutionary pressure to be stable in the presence of P69 subtilases, but proteins from more distantly-related species are not. Furthermore, in contrast to conventional resistance breeding, genetically engineering plants for human consumption is an unpopular approach with consumers, but consensus in and support from the scientific community is clear (Roberts 2018). To be able to engineer apoplasts to resemble hostile environments for plant pathogens to increase plant immunity in future, characterising apoplastic immunity landscapes of different plant species to identify novel proteins involved in plant immunity will be crucial.

4 AgromonasLux

4.1 Introduction

With the advance of high-throughput genomics and proteomics data, implicating an array of proteins in plant immunity has become feasible for biologists. Providing experimental evidence in disease assays to prove their implication in plant immunity has remained a low-throughput, time-consuming endeavour, limiting a more rapid advance of our understanding of plant immune systems. Two central limiting factors have slowed down studying proteins in plant immunity: First, laborious generation of transgenic lines in which proteins of interest are either depleted or over expressed and second, time-consuming assessment of effects of these proteins on pathogen growth. In bacterial disease assays, effects of over expressed or depleted proteins are often analysed by determining bacterial titers through colony counts. While colony counts have been the standard in studying bacterial plant pathogens, they are time consuming and limit higher throughput.

In recent years, disease assays have been developed and optimised to improve scalability when studying bacterial plant pathogens. The first bottleneck of generating stable transgenic lines was addressed by the agromonas disease assay (Buscaill et al. 2021). In agromonas, silencing and over expression of immunity-related genes via *Agrobacterium tumefaciens*-mediated transient expression (agroinfiltration) was combined with subsequent infection with *Pseudomonas syringae* (Buscaill et al. 2021). A combination of different antibiotics facilitated the selective plating of *Pseudomonas syringae* to enable colony counts and analyse effects of immunity-related proteins on bacterial titers in agroinfiltrated leaves (Buscaill et al. 2021). While agromonas overcame the limiting factor of slow generation of stable transgenic lines, the colony count-based read out remained a bottleneck for higher throughput in agromonas. The authors discussed future alternatives to increase the throughput of agromonas, including replacing colony counts with bioluminescent read outs (Buscaill et al. 2021). A concern raised was that bioluminescence assesses the bacterial metabolic state which might deviate from bacterial titers derived from colony counting methods. For years, bioluminescence-based read outs for plant defence responses

have been developed involving stable transformation of the *luxCDABE* operon under control of a constitutive promoter into the bacterial chromosome (Fan et al. 2008; Soldan et al. 2020; Jutras et al. 2021; Matsumoto et al. 2022). Bioluminescent *P.syringae* pathovars have been generated to infect french bean (*Phaseolus vulgaris*), *Arabidopsis thaliana, Solanum lycopersicum, N.benthamiana* and *Marchantia polymorhpa* (Fan et al. 2008; Soldan et al. 2020; Matsumoto et al. 2022). Strong correlation was demonstrated between bioluminescence and colony count-determined titers for *in vitro* bacterial cultures (Soldan et al. 2020), time courses *in planta* (Fan et al. 2008; Matsumoto et al. 2022) and plant defence responses (Fan et al. 2008; Matsumoto et al. 2022). While bioluminescence is also dependent on the metabolic state of the bacterium, it is a strong proxy to evaluate bacterial titers *in planta*. All described papers with bioluminescent *P.syringae* had in common that they relied on stable knock outs of either plants or pathogens to assess the effects of plant defence responses on bacterial bioluminescence. In this chapter, we aimed to combine recent research to expand agromonas by bioluminescent read out to establish the disease assay called 'agromonasLux'. AgromonasLux combines transient expression of immunity-related genes with bioluminescent read out of effects on *P.syringae* pv. *tomato* DC3000 ΔhopQ (*Pto*). Therefore, *Pto* was stably transformed using mini-Tn7-*lux* developed for Soldan et al. 2020 and Jutras et al. 2021. In this chapter, we established the experimental conditions to assess the impact of transiently over expressed plant defence proteins on bioluminescence of *Pto*. The newly-established assay was then used to test proteins implicated in plant immunity from **chapter 2** for effects on *Pto* bioluminescence.

4.2 Materials and Methods

4.2.1 Molecular cloning

Coding sequences of proteins were extracted from the Solgenomics ITAG4.0 annotation and sequences were codon-optimised to use common *N.benthamiana* codons with Geneious prime v2022.0.1 and the kazusa codon usage table of *N.benthamina* (Geneious 2022). If SignalP5.0 predicted a protein to be secreted, we replaced the SignalP5.0-predicted signal peptide with the NtPR1 signal peptide (ATGGGATTTGTTCTCTTTTCACAATTGCCT TCATTTCTTCTTGTCTCTACACTTCTCTTATTCCTAGTAATATCCCACTCTTGCCGTGCAGG AGGT) and if SignalP5.0 did not predict a protein to be secreted, but TargetP2.0 did, we replaced the TargetP2.0-predicted signal peptide with the NtPR1 signal peptide. We furthermore added a glycine-glycine linker to the 3' end of the NtPR1 signal peptide and a glycine-glycine linked 6xHis-tag (GGGGGACATCACCATCACCATCACTAA) to the 3' prime end of the codon-optimsed CDSs without their native stop codons. Furthermore, we added 5' (GAAGACATA) and 3' (GCTTATGTCTTC) BpiI restriction sites generating matching overhangs for the target vector pJK187 (pL0V2-2x35S::SC). The target vector pJK187 was based on pJK001 (Paulus et al. 2020), pAGM4723, pICH41308, pICH51288 and pICH41414 (Weber et al. 2011; Engler et al. 2014) and contained a double cauliflower mosaic virus 35S promoter and 35S terminator, a *nptII* Kanamycin resistance gene and *lacZ* as the fragment to be replaced by restriction digestion and ligation. Synthesised gene fragments are summarised in **Table 8**. Large genes were ordered in two fragments

Table 8: Tomato single gene fragments for cloning

ID	Name	First mature codon position	Predictor	Fragment ID
Solyc01g105070.3.1	Peroxidase 25-like	32	TargetP2.0	C22
Solyc02g022850.1.1	Berberine bridge-like 19	21	SignalP5.0	C23
Solyc02g064940.1.1	Extensine-like LRR	20	SignalP5.0	C24
Solyc02g077040.4.1	Pip1	30	TargetP2.0	C25
Solyc03g020030.3.1	I20 protease inhibitor	26	SignalP5.0	C26
Solyc06g083020.3.1	SCPL[1]	27	SignalP5.0	C29
Solyc07g045440.1.1	FAP1	26	SignalP5.0	C32
Solyc08g066810.3.1	GH5[2]	42	TargetP2.0	C33
Solyc09g098540.3.1	Chitinase class I	22	SignalP5.0	C34
Solyc12g009650.2.1	Proline-rich LTP[3]	26	SignalP5.0	C35

1: SCPL: Serine carboxypeptidase-like, 2: GH: Glycosyl hydrolase 5, 3: LTP: Lipid transfer protein.

with BpiI restriction sites producing unique overhangs to combine both fragments during the BpiI - T4 DNA ligase digestion-ligation reaction. Mutations were introduced into P69B by varying one of the gene fragments to create a modular system. Two-component gene fragments are described in **Table 9**. Gene fragments were resuspended in 10 mM

Table 9: Tomato multi fragment genes

ID	Name	First codon	Last codon	Mutations	Predictor	Fragment ID
Solyc08g007670.1.1	SBT1.10q[1]	22	406	-	SignalP5.0	C36
Solyc08g007670.1.1	SBT1.10q[1]	406	757	-	-	C37
Solyc08g079870.3.1	P69B	23	425	-	SignalP5.0	C38
Solyc08g079870.3.1	P69B	23	425	L398F, H400D	SignalP5.0	C39
Solyc08g079870.3.1	P69B	23	425	L398F, H400G	SignalP5.0	C40
Solyc08g079870.3.1	P69B	23	425	L398F, H400R	SignalP5.0	C41
Solyc08g079870.3.1	P69B	425	758	-	-	C42

1: The Bpi1 restriction site is within this codon.

Tris-HCl, pH 8.0 to a final concentration of 40 ng/µl. To clone gene fragments into the pJK187 plasmid we mixed 50 ng of pJK187 with 40 ng of the gene fragments in 0.5 µl BSA (10 mg/ml), 0.5 µl 10x T4 DNA ligase buffer (Thermo Scientific, Waltham, USA), 0.25 µl (1.25 U) T4 DNA ligase (Thermo Scientific, Waltham, USA), 0.25 µl (2.5U) BpiI (Thermo Scientific, Waltham, USA) and sterile milliQ water to a final volume of 5 µl. Solutions were incubated for five cycles, each 20 minute cycle consisting of 10 minutes at 37°C and 10 minutes at 16°C, followed by a final cycle of 10 minutes at 37°C and 20 minutes at 80°C. 2.5 µl of the digestion-ligation reaction were added to 50 µl competent DH10β *E.coli* and kept on ice for 30 minutes, followed by 45 seconds at 42°C and 2 minutes on ice. Then, 300 µl LB medium was added and cell suspensions incubated at 37°C for two hours and then plated on LB plates with 50 µg/ml kanamycin, 0.1 mM IPTG, 20 µg/ml X-Gal. Plates were incubated over night at 37°C. Single white colonies were used to inoculate 5 ml LB medium liquid cultures to grow over night at 37°C. Plasmids were isolated with the QIAgen Spin Miniprep kit and eluted in 40 µl elution buffer. Plasmids were sequenced using Source Bioscience using 35S promoter (CTATCCTTCGCAAGACCCTTC) and terminator (CTCAACACATGAGCGAAACC) primers to confirm the inserts. Validated binary plasmids were transformed into *A.tumefaciens* GV3101 pMP90 via heat shock transformation. 1 µg of plasmid was mixed with 20 µl

competent *A.tumefaciens* GV3101 and incubated on ice for 30 minutes, followed by 2 minutes in liquid nitrogen and 5 minutes at 37°C before adding 600 µl of LB medium and incubating at 28°C. After three hours, 100-200 µl of the cell suspension was plated onto LB plates containing 50 µg/ml kanamycin and gentamycin each, and 25 µg/ml rifampicin and incubated at 28°C over night.

All plasmids used in this chapter are described in **table 10**.

Table 10: ALl plasmid used in this chapter

Plasmid	Description	Reference
pFH05	Binary plasmid for peroxidase-His (Solyc01g105070.3.1) expression	This thesis
pFH06	Binary plasmid for Berberine bridge-like-His (Solyc02g022850.1.1) expression	This thesis
pFH07	Binary plasmid for Extensie-like LRR-His (Solyc02g064940.1.1) expression	This thesis
pFH08	Binary plasmid for Pip1-His (Solyc02g077040.4.1) expression	This thesis
pFH09	Binary plasmid for I20 protease inh.-His (Solyc03g020030.3.1) expression	This thesis
pFH15	Binary plasmid for FAP1-His (Solyc07g045440.1.1) expression	This thesis
pFH16	Binary plasmid for GH5-His (Solyc08g066810.3.1) expression	This thesis
pFH17	Binary plasmid for Chitinase class I-His (Solyc09g098540.3.1) expression	This thesis
pFH18	Binary plasmid for proline-rich, LTP-His (Solyc12g009650.2.1) expression	This thesis
pFH19	Binary plasmid for SBT1.10q-His (Solyc08g007670.1.1) expression	This thesis
pFH20	Binary plasmid for P69B-His (Solyc08g079870.3.1) expression	This thesis
pFH21	Binary plasmid for P69B_L398F_H400D-His expression	This thesis
pFH22	Binary plasmid for P69B_L398F_H400G-His expression	This thesis
pFH23	Binary plasmid for P69B_L398F_H400R-His expression	This thesis
pJK001c	Empty vector controls	(Paulus et al. 2020)
pJK668	Binary plasmid for *Sl*FLS3 (Solyc04g009640) expression	(Buscaill et al. 2021)
pPB046	Binary plasmid for *Nb*PRP27 expression	(Morimoto et al. 2022)
pAG001	Binary plasmid for *Nb*PR3 expression	(Godson et al., *in preparation*)
pPJ048	Binary plasmid for cytoplasmic GFP expression	(Jutras et al. 2021)
pPJ057	Binary plasmid for secreted RFP expression	(Jutras et al., *unpublished*)

4.2.2 AgromonasLux

The workflow was predominantly based on the agromonas assay (Buscaill et al. 2021). Fresh *A.tumefaciens* colonies were picked separately from LB plates to inoculate 10 ml of LB medium with respective antibiotics and incubated shaking at 200 rpm, 20°C for 16-20

hours. Then, bacterial cultures were centrifugated at 4000 x g, 20°C for 2 minutes, and the bacterial pellet was resuspended to OD_{600} of 0.5 in infiltration buffer (10 mM MES, 10 mM $MgCl_2$ at pH 5). *A.tumefaciens* containing a cGFP and p19 encoding plasmid were mixed at equal ratio with *A.tumefaciens* transformed with plasmids encoding genes of interest to a final OD_{600} of 0.5. Whole 3 week-old *N.benthamiana* leaves were syringe-infiltrated after poking small wholes on the abaxial infiltration site of leaves using a needle. Fresh *P.syringae* pv. *tomato* DC3000 ΔhopQ (Wei et al. 2007) colonies were picked from LB plates to inoculate 10 ml LB medium without antibiotics as overnight cultures, incubating at 200 rpm at 28°C. After 15-20 hours, 200 µl of over night cultures were used to inoculate fresh growth cultures in 10 ml LB medium without antibiotics, incubating at 200 rpm and 28°C for 3-4 hours. Then, *Pto* growth cultures were centrifuged at 4000 x g, 20°C for 2 minutes. The bacterial pellet was resuspended in sterile milliQ water and the OD_{600} was adjusted by 1:10 dilution series starting from $OD_{600} = 0.1$. Previously agroinfiltrated leaves were syringe infiltrated to infect with *Pto* after poking small wholes on the abaxial infiltration site of leaves using a needle. Previous sites of agroinfiltration and areas directly next to large veins were avoided. Single *Pto* infiltration covered areas large enough to sample at least one 6 mm diameter leaf disc away from direct infiltration sites and large leaf veins.

4.2.3 Apoplastic fluid extraction

To extract apoplastic fluid, we infiltrated *N.benthamiana* leaves with ice-cold sterile milliQ water using a disposable 50 ml syringe (O'Leary et al. 2014). After infiltrating leaves, we dried their surfaces and rolled them up carefully to fit into disposable 20 ml syringes. Each syringe was placed into a sterile 50 ml falcon tube and stored on ice until samples were ready for centrifugation. Next, we extracted apoplastic fluid from infiltrated leaves by centrifugation for 25 minutes, at 4°C and 500 x g. After extraction, apoplastic fluid was aliquoted and flash frozen in liquid nitrogen as quickly as possible.

4.2.4 Western Blot analysis

20 µl apoplastic fluid extract were mixed with 5 µl protein gel loading buffer (200 mM Tris-HCl, pH 6.8, 400 mM DTT, 8% SDS, 0.4% bromophenol blue and 40% glycerol) and incubated at 95°C for 5 minutes. Apoplastic fluid samples were then separated by SDS-PAGE on 12-18% gels. Proteins were transferred onto BioRad nitrocellulose membranes using the BioRad Trans-Blot Turbo Transfer system. Afterwards, the membrane was blocked with 1x PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM $NaHPO_4$, 1.8 mM KH_2PO_4) with 5% nonfat, dry milk powder for one hour at RT. Then, the membrane was washed in 20 ml PBST (1x PBS buffer with 0.05% Tween-20) twice for 5 minutes. Afterwards, the membrane was incubated in 1x PBS with 5% nonfat, dry milk powder and 1:5000 ratio monoclonal anti C-terminal 6x-His antibody, conjugated with horse radish peroxidase (invitrogen #46-0707) for 2 hours. Then, the membrane was washed twice in PBST for 5 minutes. 1 ml of Clarity Western ECL Substrate (BioRad) was added and chemiluminescence detected using ImageQuant LAS 4000 (GE Healthcare).

4.2.5 Plate reader luminescence read out

Leaf discs were taken with a cork borer (diameter 6 mm) and placed abaxial side facing up on 200 µl milliQ water in white 96 well plates. Leaf discs were stored in the dark for 10 minutes in the Tecan infinite 200 pro before measuring luminescence with 10.000 ms integration time. We measured cytoplasmic GFP at λ(excitation)= 500 nm and λ(emission)=530 nm with gain 50 and secreted RFP was detected at λ(excitation)=550 nm and λ(emission)=585 nm with gain 50.

4.2.6 Statistics

All data points of biological replicates are means of 3-4 samples from separate infection events of the same plant. Statistical comparisons included one-way analysis of variance (ANOVA) and block design was added if experiments from multiple dates were combined. We corrected one-way ANOVA-derived p-values for multiple testing using the Benjamini-Hochberg procedure with a significance threshold of 0.05.

4.3 Results

The aim of this chapter was to establish a bioluminescence-based readout for the agromonas infection assay of *N.benthamiana* to test tomato apoplastic proteins from **chapter 2** for their implication in plant immunity. Bioluminescent *Pseudomonas syringae* pv. *tomato* DC3000 ΔhopQ (*Pto*) was generated by Riccardo Soldan to enable luminescence-based read out of bacterial density. The new bioluminescent agromonas infection assay will be referred to as 'agromonasLux'.

4.3.1 Experimental design

To monitor bacterial growth in agroinfiltrated leaves, we defined the experimental setup to integrate bioluminescent *Pto* in the Agromonas disease assay (Buscaill et al. 2021). Each tray included five 3-week old plants from different treatments and one plant was used for a single treatment. An experiment with five treatments and eight biological replicates each thus included eight trays. Between trays, plants of different treatments were rotated to reduce bias due to the position in a tray (**Figure 39A**). For each plant, two true leaves with a diameter of 3-5 cm were used per treatment (**Figure 39B**). Whole leaves were agroinfiltrated and each leaf was infiltrated with bioluminescent *Pto* in two separate areas. Each *Pto*-infected area was sampled once per time point, resulting in four pseudoreplicates per plant (**Figure 39B**). Pseudoreplicates are uniquely-correlated biological replicates taken from the same individual and while these multiple measurements describe one individual, they are not independent descriptions of the individual's population as a whole (Hurlbert 1984; Blainey et al. 2014). Pseudoreplicates were thus averaged to more robustly determine the mean for each plant. Each plant is considered a single biological replicate. The assay was developed in close collaboration with Pierre Buscaill and we adopted conditions defined in agromonas, such as plant age at agroinfiltration (3 weeks), time points of agroinfiltration (0 dpi), Pto infection (2 dpi) and imaging (5 dpi) and density of *A.tumefaciens* infiltration cultures (final $OD_{600} = 0.5$) from Buscaill et al. 2021.

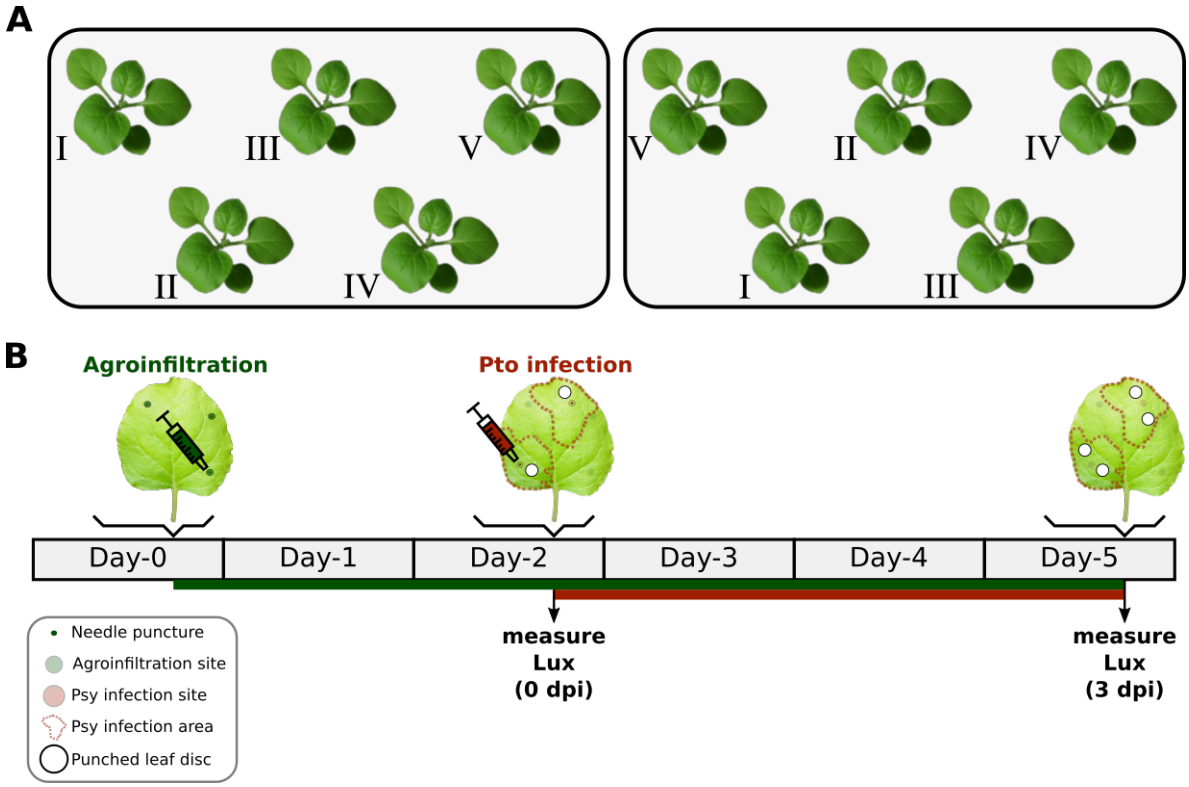


Figure 39: AgromonasLux - general experimental design. (**A**) Layout of two plant trays with each five 3-4 week-old *N.benthamiana* plants for the AgromonasLux assay. Images of plants were modified from (Kamoun 2022). The tray contained all treatments (I to V) from the experiment and differently-treated plants were rotated between trays. (**B**) Design of agroinfiltration and *P.seudomonas syringae* pv. *tomato* DC3000 ΔHopQ (*Pto*) infection using an adult *N.benthamiana* leaf. The design of the workflow was adopted from Buscaill et al. 2021. For the experiments, leaves were not detached and for each plant, two true, not fully developed leaves with a diameter of approximately 3-5 cm were infiltrated. Whole leaves were agroinfiltrated with a syringe and two days later, the same leaves were infected with bioluminescent *Pto* at two different sites. The two sites were at opposing ends of the length and width of the leaves. *Pto*-infection areas were mostly non-overlapping and large enough to sample leaf discs (6 mm diameter) away from the infection site and major veins. The image of the *N.benthamiana* leaf was modified from Wikipedia.

4.3.2 Bioluminescent *Pto*

The mini-Tn7-*lux* vectors pRS-p$_{oxB13}$::lux and pRS-p$_{nptII}$::lux were used to generate *Pto* with weak (*oxb13*) and strong (*nptII*) constitutive promoters driving *luxCDABE*-mediated bioluminescence (Soldan et al. 2020). The *oxb13* promoter is based on the *recA* promoter and has been described in *Pseudomonas syringae* pv. *phaseolicola* before (Soldan et al. 2020). In the following, Pto pRS-p$_{oxb13}$::lux and pRS-p$_{nptII}$::lux transformants will be referred to as *Pto(oxb13:lux)* and *Pto(nptII:lux)* respectively. To determine dif-

ferences in luminescence between *Pto(nptII:lux)* and *Pto(oxb13:lux)*, we grew each strain in liquid culture and measured luminescence at different OD_{600} of 1:10 dilution series (**Figure 40**). Since luminescence of *Pto(oxb13:lux)* was close to the detection limit at

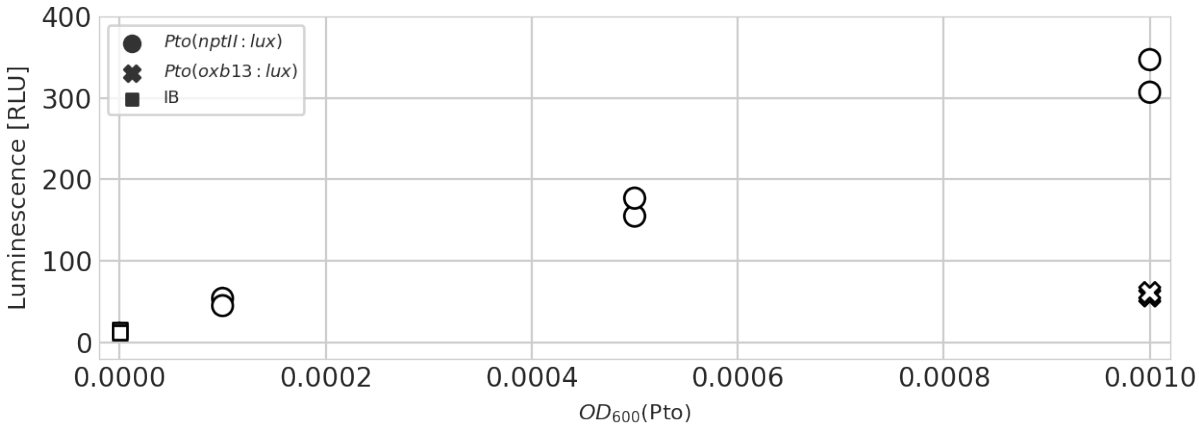


Figure 40: *Pto(nptii:lux)* shows higher luminescence *in vitro* compared to *Pto(oxb13:lux).* Two technical replicates of bioluminescence at different bacterial densities of *Pto* in infiltration buffer (IB). IB represents infiltration buffer without added *Pto(oxb13:lux)* or *Pto(nptII:lux).*

OD_{600}=0.0010 already, we did not measure measure luminescence at higher dilutions. At OD_{600}=0.0010, bacterial cultures of *Pto(nptII:lux)* show 5.5-fold higher luminescence compared to *Pto(oxb13:lux)*, which is in line with literature for *P.syringae* pv. *phaseolicola* 1302A (Soldan et al. 2020). Furthermore, luminescence of *Pto(nptII:lux)* infection cultures was detectable at low bacterial densities of OD_{600} = 0.0005 and 0.0001.

To compare detection of bioluminescent *Pto in planta*, we infected agroinfiltrated leaves with bioluminescent *Pto* at the same OD_{600} as for the *in vitro* experiment and measured luminescence in leaf discs at 0 and 3 dpi (**Figure 41**). *Pto(oxb13:lux)* had not been tested at OD_{600} <0.0010, as cultures were close to the luminescence detection limit *in vitro* at OD_{600}=0.0010, which made lux-based verification of infection densities at lower OD_{600} unreliable. Directly after infiltration (0 dpi), luminescence of *Pto(oxb13:lux)* and *Pto(nptII:lux)* were indistinguishable (FDR multiple testing-corrected p-value of one-way ANOVA: 0.50) (**Figure 41A**). At infection OD_{600}, luminescence strongly increased for both *Pto(oxb13:lux)* and *Pto(nptII:lux)* infected leaves after three days and was 8.5-fold higher in *Pto(nptII:lux)* than *Pto(oxb13:lux)* (**Figure 41B**). Furthermore, we detected

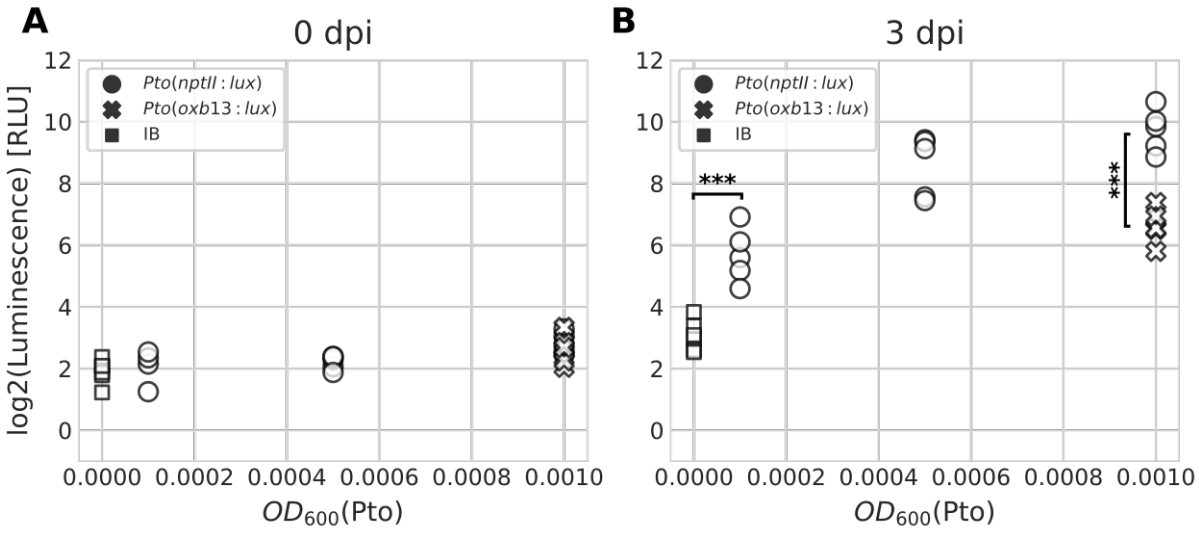


Figure 41: *Pto(nptii:lux)* shows higher luminescence in agroinfiltrated *N.benthamiana* leaves compared to *Pto(oxb13:lux)* after 3 days. Agroinfiltrated *N.benthamiana* leaves expressing EV infected with *Pto(nptII:lux)*, *Pto(oxb13:lux)* and infiltration buffer-only (IB) For IB, no bacteria had been added. Each data point represents the mean of 3-4 measurements from different infection events of the same plant all measurements are from the same experiment. All n=5 except n(IB)=6. Benjamini-Hochberg, false discovery rate-corrected p-values of one-way ANOVAs are indicated as: $***< 0.001$.

a significant increase in luminescence for *Pto(nptII:lux)* at OD_{600}=0.0001 relative to the mock-infected (IB) control. Based on its higher sensitivity, only *Pto(nptII:lux)* was used in following experiments and due to low luminescence signal at 0 dpi, no further measurements were taken directly after infection to reduce mechanical damage to plants.

4.3.3 Effects of immune proteins on bioluminescence of *Pto in planta*

To test if we can detect apoplastic immunity-mediated reduction in luminescence, we over expressed immunity proteins *Sl*FLS3, *Nb*PR3 and *Nb*PRp27 in plants mediated by agroinfiltration and we analysed resulting effects on luminescence at different *Pto* infection densities. Recombinant proteins were over expressed by the plant, mediated by *A.tumefaciens*. *Solanum lycopersicum* FLAGELLIN-SENSING 3 (*Sl*FLS3) encodes a receptor-like kinase conferring moderate resistance to *Pto* and transient over expression confers resistance to *Pto* in *N.benthamiana* (Hind et al. 2016; Buscaill et al. 2021). *N.benthamiana* PRp27 (*Nb*PRp27) encodes a metalloprotein contributing to resistance

to *P.syringae* pv. *tabaci* and pv. *tomato* (Xie and Goodwin 2009; Morimoto et al. 2022). *N.benthamiana* PR3 encodes a class I chitinase-like protein contributing to resistance against *Pto* [Godson et al., (textitin preparation). All three proteins increase resistance to *Pto* in agromonas assays (Buscaill et al. 2021; Morimoto et al. 2022, Godson et al., *in preparation*). Agroinfiltrated leaves expressing empty vector (EV) were used as negative controls and all constructs were co-infiltrated with *A.tumefaciens* transformed with a plasmid to mediate expression of cytoplasmic GFP (cGFP) and silencing inhibitor p19 by the plant. We compared *Pto(nptII:lux)* infection densities $OD_{600} = 0.0001$ and $OD_{600} = 0.0010$ and hypothesised that immunity-related effects would be stronger at lower infection densities. Agroinfiltrated leaves expressing *Sl*FLS3 showed significantly reduced luminescence after three days, but agroinfiltrated leaves expressing *Nb*PR3 and *Nb*PRp27 did not (FDR-corrected p-values of 0.24 and 0.70 respectively) (**Figure 42A**). Evidence of recombinant protein expression in agroinfiltrated leaves was provided by GFP fluorescence (**Figure 42A**). GFP fluorescence in agroinfiltrated leaves expressing *Sl*FLS3 and *Nb*PRp27 was reduced compared to EV controls (**Figure 42A**), which could be due to competition for cellular resources or transcriptional silencing resulting in reduced cGFP production. In contrast, transiently over expressing *Nb*PR3 showed strongly induced increase in GFP fluorescence (**Figure 42A**). This indicated that cytoplasmic protein levels likely increased strongly upon *Nb*PR3 expression, despite no detectable difference in luminescence. It was later found, however, that this strain resulting in PR3 expression by the plant also resulted in agrobacterium-mediated expression of cGFP (Godson, *unpublished*). Next, we tested effects of transient expression of *Sl*FLS3 and *Nb*PR3 at higher *Pto(nptII:lux)* infection levels. Due to time limitations, we did not test *Nb*PRp27 at *Pto* infection $OD_{600} = 0.0010$. At $OD_{600} = 0.0010$, we observed significant differences for both proteins relative to EV controls (**Figure 43**). Over expressing *Sl*FLS3 showed similar reduction in luminescence at both high and low infection densities (3.0-fold and 2.7-fold reduction, respectively). Fluorescence of cGFP showed again evidence of recombinant protein expression in agroinfiltrated leaves, but this time, we did not observe a significant decrease in GFP fluorescence in leaves expressing *Sl*FLS3, whereas GFP fluorescence strongly in-

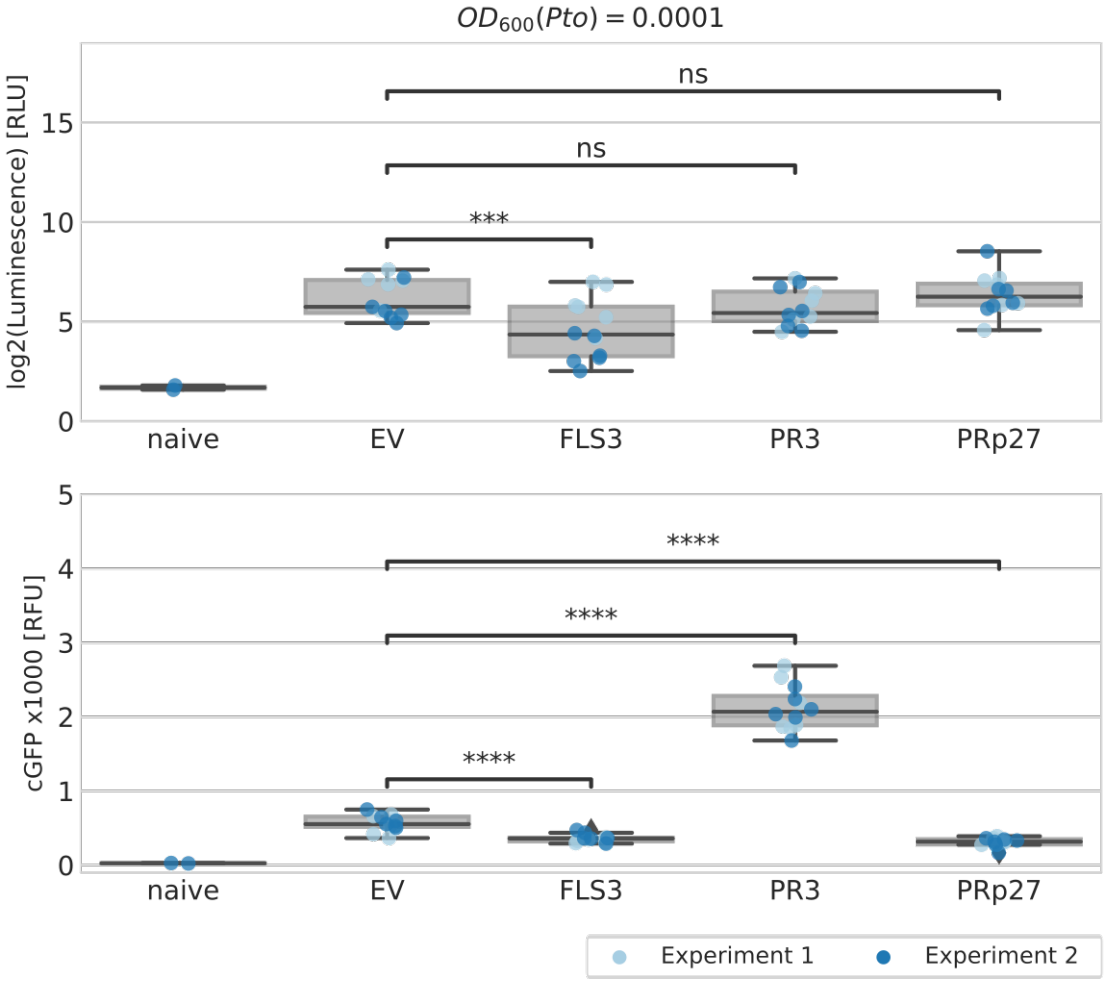


Figure 42: AgromonasLux detects apoplastic immunity-mediated reduction of bacterial luminescence. Luminescence and fluorescence of *Pti(nptII:lux* infected and agroinfiltrated leaves expressing apoplastic immunity components. All constructs, empty vector (EV), *Sl*FLS3 (FLS3), *Nb*PR3 (PR3) and *Nb*PRp27 (PRp27) were co expressed with cytoplasmic GFP and p19. Naive leaf discs represent untreated leaf discs. Luminescence and fluorescence are expressed as relative luminescence units (RLU) and relative fluorescence units (RFU) respectively. Luminescence measures bacterial density and cGFP (excitation 500 nm, detection 530 nm) is a representing cytoplasmic recombinant protein production in leaf discs. Data points represent the means of three to four samples from the same plant and colours indicate the respective date of the experiment. For all comparisons, biological replicates were between 10 and 12 and comprised of two experimental replicates. Centre lines are the median, bounds of boxes the 25th and 75th percentiles and whiskers 1.5 x interquartile range (IQR) from 25th and 75th percentiles, outliers are indicated as diamonds. Statistical testing was based on one-way blocked ANOVA, using the date of experiments as blocks. Benjamini-Hochberg, false discovery rate-corrected p-values are indicated as: ns> 0.05, $***< 0.001$, $****< 0.0001$.

creased for over expression of *Nb*PR3 (**Figure 43**). We concluded that at *Pto* infection $OD_{600} = 0.0010$, we robustly detect reduced bioluminescence of *Pto(nptII:lux)* in agroinfiltrated leaves over expressing apoplastic immunity proteins *Sl*FLS3 and *Nb*PR3.

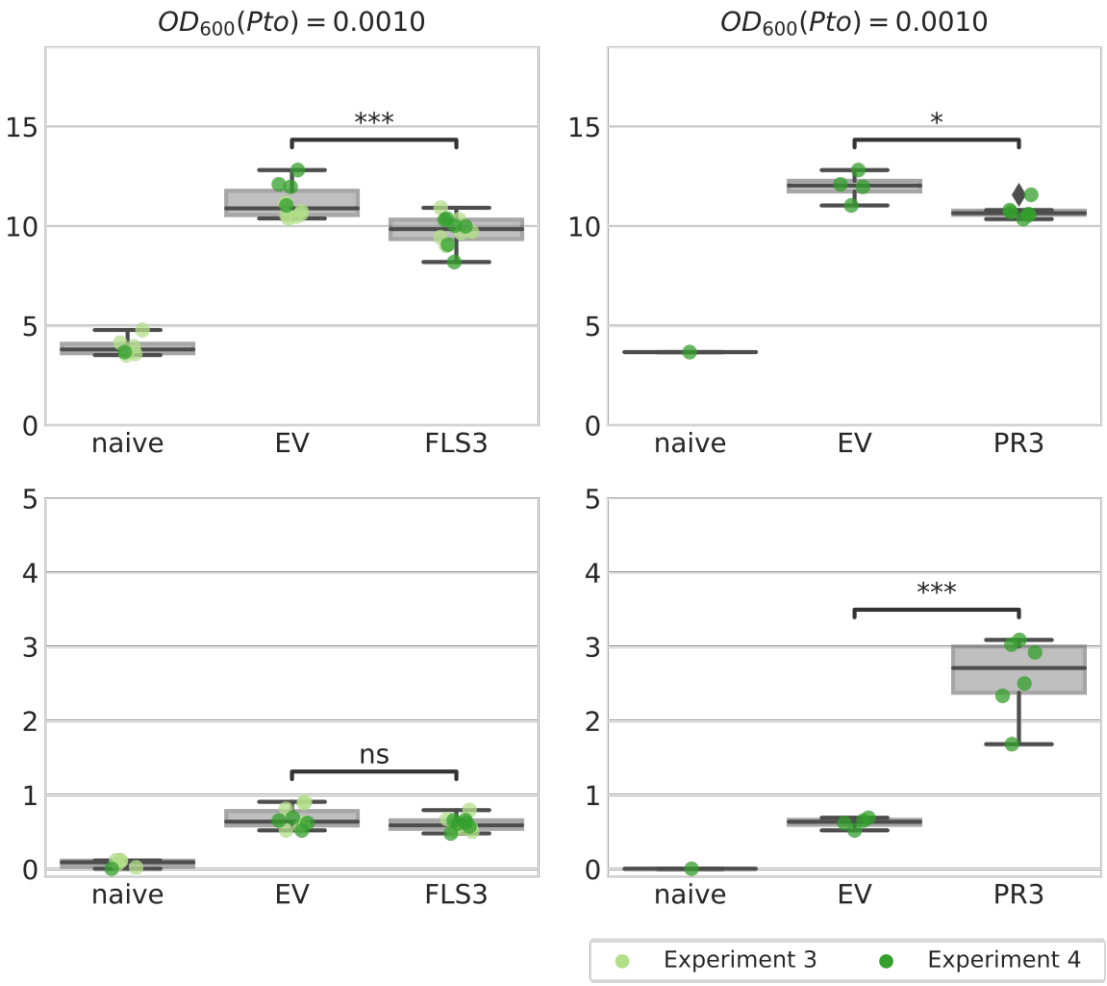


Figure 43: AgromonasLux detects apoplastic immunity-mediated reduction of bacterial luminescence. Luminescence and fluorescence of *Pto(nptII:lux)*-infected and agroinfiltrated leaves expressing apoplastic immunity components. All constructs, empty vector (EV), *Sl*FLS3 (FLS3), *Nb*PR3 (PR3) and *Nb*PRp27 (PRp27) were co expressed with cytoplasmic GFP and p19. Naive leaf discs represent untreated leaf discs. Luminescence and fluorescence are expressed as relative luminescence units (RLU) and relative fluorescence units (RFU) respectively. Data points represent the means of three to four samples from the same plant and colours indicate the respective date of the experiment. For EV-FLS3 comparisons: n(EV)=10 and n(FLS3)=12 and for EV-PR3 comparison, n(EV)=4 and n(PR3)=6. Centre lines are the median, bounds of boxes the 25th and 75th percentiles and whiskers 1.5 x interquartile range (IQR) from 25th and 75th percentiles, outliers are indicated as diamonds. Statistical testing was based on one-way blocked ANOVA, using the date of experiments as blocks. Benjamini-Hochberg, false discovery rate-corrected p-values are indicated as: ns> 0.05, $* < 0.05$, $*** < 0.001$, $**** < 0.0001$.

Because empty vector is not a condition that is comparable to secreted (PR) proteins, we decided to include secreted RFP as a negative control. We therefore included both negative controls (EV and secRFP) and positive controls (FLS3, PR3) at medium infection density ($OD_{600} = 0.0010$). As before at three days post *Pto* infection, luminescence was

reduced in agroinfiltrated leaves expressing *Sl*FLS3 and *Nb*PR3 relative to EV control (**Figure 44**). Compared to secRFP, luminescence was also reduced by *Sl*FLS3, but not

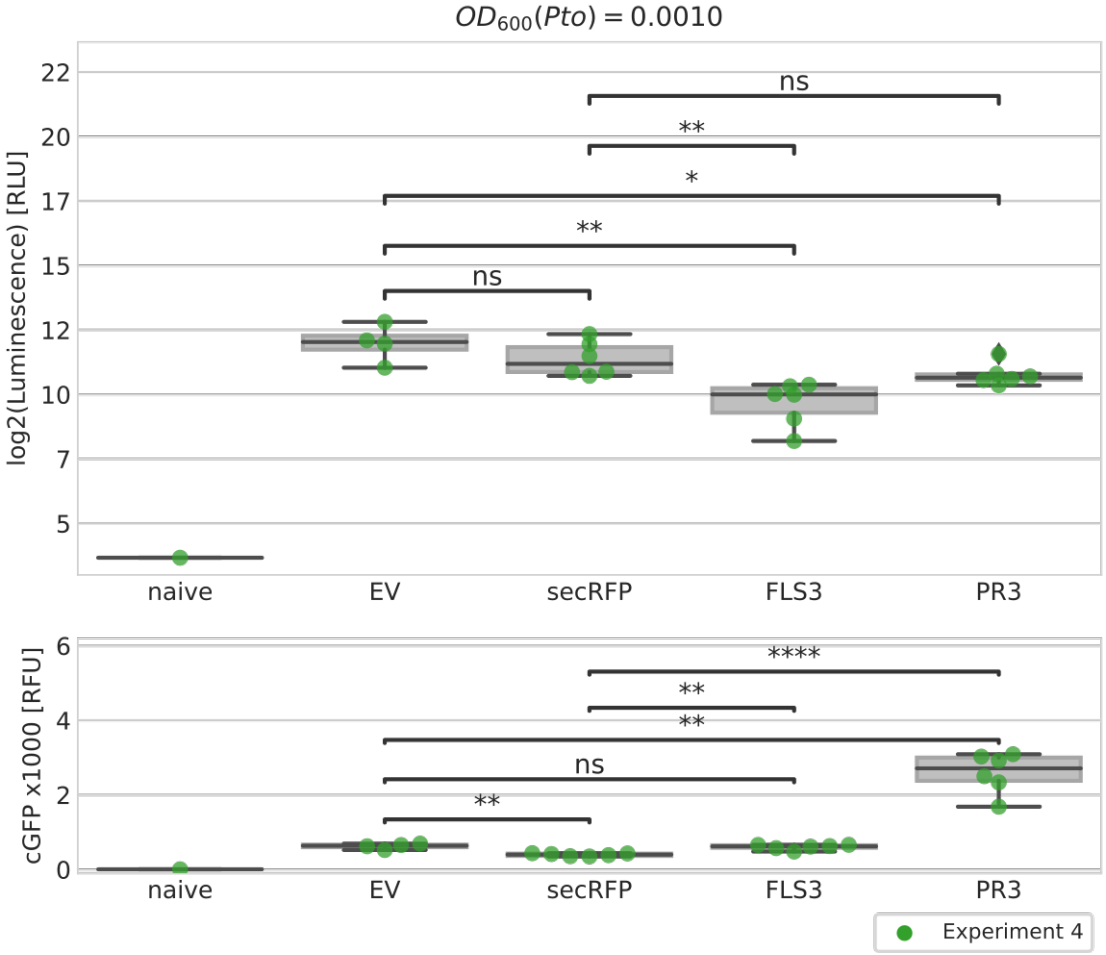


Figure 44: Comparison of different negative controls for secreted immunity-related proteins. Luminescence and fluorescence are expressed as relative luminescence units (RLU) and relative fluorescence units (RFU) respectively. This data set shows all biological replicates of experiment 4. All constructs, empty vector (EV), secreted RFP (secRFP), *Sl*FLS3 (FLS3) and *Nb*PR3 (PR3) were co expressed with cytoplasmic GFP and p19. Naive leaf discs represent untreated leaf discs. Data points represent the means of three to four samples from the same plant and colours indicate the respective date of the experiment. For all comparisons, n=[4, 6]. Centre lines are the median, bounds of boxes the 25th and 75th percentiles and whiskers 1.5 x interquartile range (IQR) from 25th and 75th percentiles. Diamonds represent outliers. Statistical testing was based on one-way blocked ANOVA, using the date of experiments as blocks. Benjamini-Hochberg, false discovery rate-corrected p-values are indicated as: ns> 0.05, $* < 0.05$, $** < 0.01$, $**** < 0.0001$.

by *Nb*PR3 (FDR-corrected p-value of 0.11) (**Figure 44**). While the difference between EV and secRFP was not statistically significant (FDR-corrected p-value of 0.21), the reduced luminescence in agroinfiltrated leaves expressing secRFP was sufficient for the

secRFP-*Nb*PR3 comparison to not be significant with six biological replicates. We therefore decided to replace EV with secRFP as a negative control and to increase the number of biological replicates from six to eight. Interestingly, secRFP expression reduced cGFP fluorescence compared to the EV control, indicating secRFP expression reduces cGFP production or accumulation (**Figure 44**).

4.3.4 Test tomato candidate genes in AgromonasLux

To test tomato genes from **chapter 2** for their implication in apoplastic immunity against *Pto*, we cloned 12 different genes and three P69B variants for transient expression in *N.benthamiana*. We selected tomato genes implicated in plant immunity from **chapter 2** and codon optimised their predicted coding sequences (CDSs) to include common *N.benthamiana* codons. Codons were optimised using Geneious Prime v2022.0.1 based on the kazusa codon usage table of *N.benthamiana* (Geneious 2022). The SignalP- or TargetP-predicted signal peptide was replaced by the NtPR1 signal peptide at the N-terminus and a 6x His-tag was added to the C-terminus. The NtPR1 signal peptide has been used for recombinant protein expression of apoplastic proteins in *N.benthamiana* before (Grosse-Holz et al. 2018). We cloned these gene strings into the pJK187 vector (pL0V2-2x35S::SC) which already contained a 35S promoter and terminator (**Figure 45**). Cloned genes, their functional annotations and names of plasmids are summarised in **Table 11**. Next, we assessed apoplastic protein accumulation upon transient expression in *N.benthamiana*. For each protein, three leaves from three different plants were agroinfiltrated and apoplastic fluids (AFs) were extracted three or four days later. AFs were

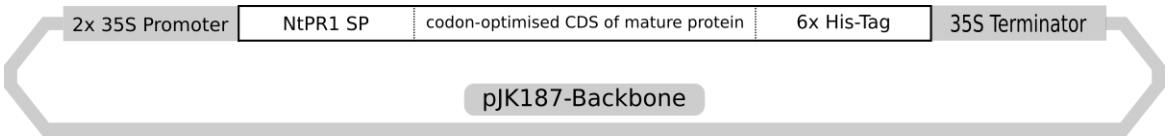


Figure 45: Cloning strategy to test tomato candidate genes in agromonasLux. Gene strings including codon-optimised coding sequence (CDS) of genes of interest, their predicted signal peptide (SP) replaced with *Nicotiana tabacum* PR1 SP (NtPR1 SP) and a 6x His-Tag were cloned into the pJK187 backbone, containing a 2x 35S promoter and 35S terminator for transient expression in *N.benthamiana*. Codon optimisation of tomato CDS was based on *N.benthamiana* codon frequency. The pJK187 plasmid was developed by Jiorgos Kourelis.

Table 11: Tomato proteins to test in AgromonasLux

ID	Functional Annotation	Name	Plasmid Name
Solyc01g105070.3.1	Class III haem peroxidase	Peroxidase 25-like	pFH05
Solyc02g022850.1.1	Berberine-bridge-like Enzyme 19	-	pFH06
Solyc02g064940.1.1	Extensine-like LRR	-	pFH07
Solyc02g077040.4.1	PLCP[1]	Pip1	pFH08
Solyc03g020030.3.1	I20 protease inh.	-	pFH09
Solyc06g083020.3.1	SCPL[2]	Serine Carboxypeptidase 37	pFH12
Solyc07g045440.1.1	Fasciclin-like arabinogalactan 1	FAP1	pFH15
Solyc08g066810.3.1	GH5[3] cellulase	-	pFH16
Solyc09g098540.3.1	Chitinase class I	-	pFH17
Solyc12g009650.2.1	Proline-rich, LTP	-	pFH18
Solyc08g007670.1.1	S8 subtilase	SBT1.10q	pFH19
Solyc08g079870.3.1	S8 subtilase	P69B	pFH20
Solyc08g079870.3.1	S8 subtilase	P69B L398F H400D	pFH21
Solyc08g079870.3.1	S8 subtilase	P69B L398F H400G	pFH22
Solyc08g079870.3.1	S8 subtilase	P69B L398F H400R	pFH23

1: PLCP: Papain-like cysteine protease. 2: SCPL: Serine Carboxypeptidase-like. 3: GH5: Glycosyl hydrolase 5.

pooled per protein and protein accumulation was analysed via SDS-PAGE and western blot using anti His-tag antibodies. No coomassie- or ponceau-stains of protein gels were imaged. We detected 7 of the 15 proteins in AF of *N.benthamiana* including a class III peroxidase (Per25, pFH05), Pip1 (pFH08), P69B (pFH20) and the three P69B variants (pFH21-23) (**Figure 46**). The remaining eight proteins were not detected which could be caused by cleavage of the C-terminal 6x His-tag or insufficient apoplastic protein accumulation. Due to time constrains, we did not further explore options to detect or express the remaining eight candidate proteins.

To test Per25, Pip1 and wild type P69B for their role in immunity, we expressed them in leaves of 3-week-old *N.benthamiana* plants via agroinfiltration. As a negative control, we used secRFP and all constructs were co-infiltrated with *A.tumefaciens* transformed with a plasmid to express cytoplasmic GFP and silencing inhibitor p19. Two days post agroinfiltration, leaves were infected with *Pto(nptII:lux)* at $OD_{600} = 0.0010$ and leaf discs for luminescence and fluorescence analysis were taken three days later. The positive control *Nb*PR3 confirmed that AgromonasLux detected antibacterial activity of apoplastic immunity proteins compared to the negative control secRFP (**Figure 47**). However, we detected no significant difference in luminescence for any of the three tested tomato proteins (**Figure 47A**). Detection of GFP fluorescence in all samples indicated that recombi-

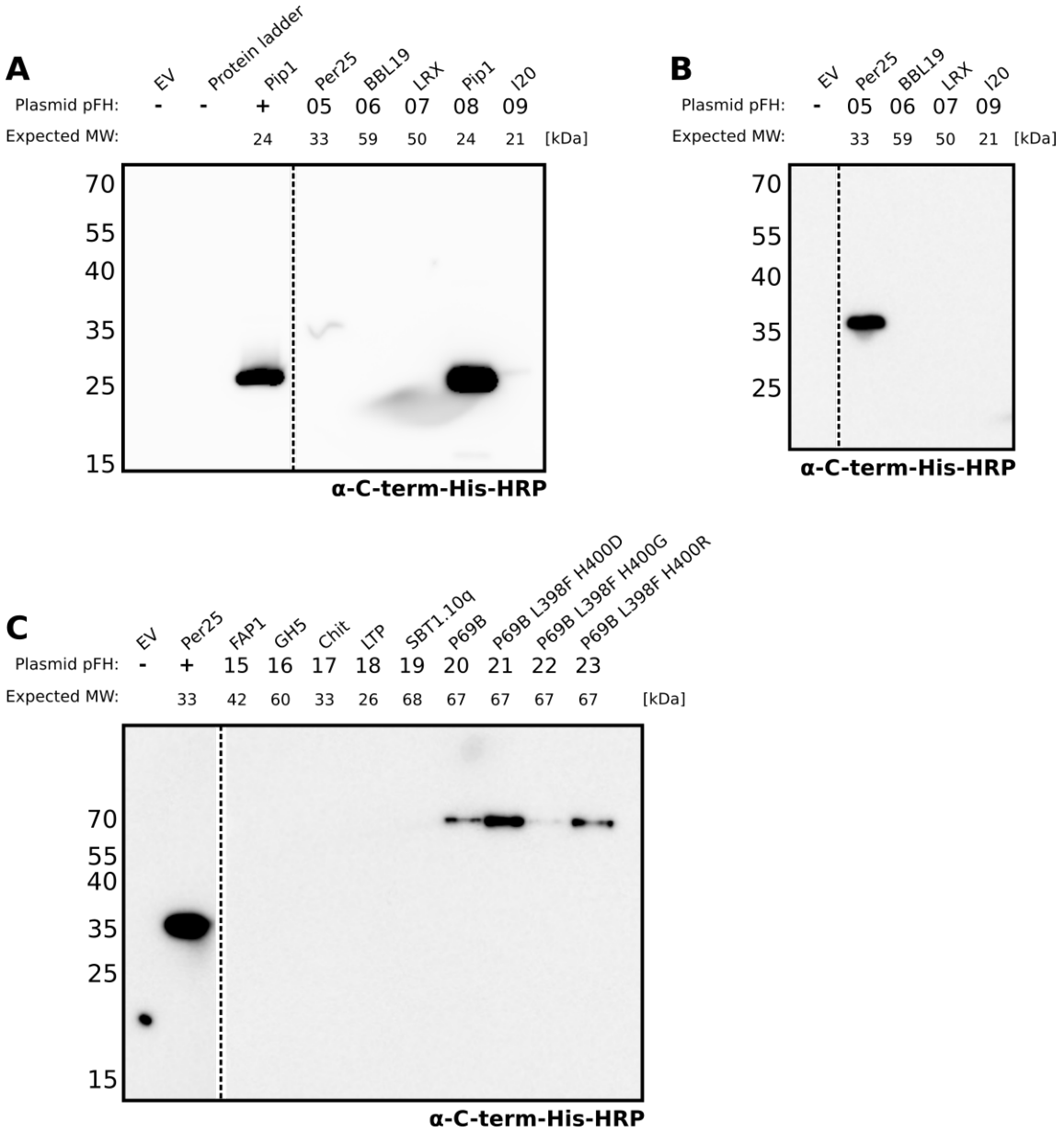


Figure 46: Candidate proteins accumulate in AF. Detection of transiently expressed, C-terminally His-tagged candidate tomato proteins in apoplastic fluid (AF) of 4.5 week-old *N.benthamiana* using anti C-terminal-His antibodies conjugated with horse radish peroxidase (α-C-term-His-HRP). (**A, B, C**) Negative controls (-) were AF from agroinfiltrated leaves expressing an empty vector and positive controls (+) were cHis-tagged Pip1 aliquots from a different construct (**A**) and Per25 from **B** in **C**. Each sample represented AFs combined from three agroinfiltrated leaves from three different plants. AF was extracted 3 (**A, B**) and 4 (**C**) days post agroinfiltration. Samples in **A** and **B** are identical and were run on separate gels and (**B**) excluded pFH08 due to its high signal masking the detection of other proteins.

nant proteins had been expressed (**Figure 47B**). For *Nb*PR3, the strongly increased GFP fluorescence was later shown to be caused by this PR3-expressing strain also expressing GFP (Godson, *unpublished*).

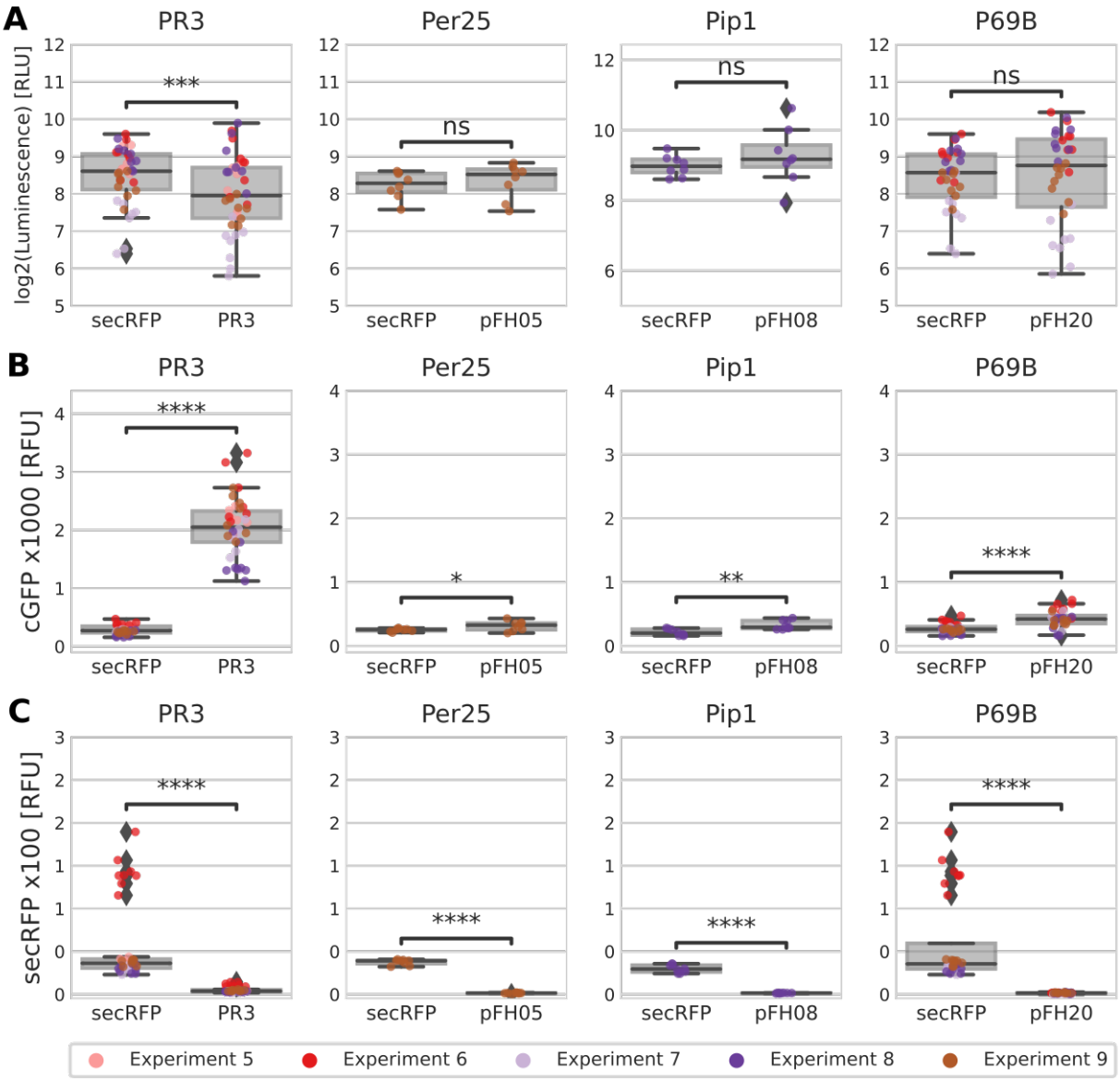


Figure 47: Tomato candidate proteins in agromonasLux. Luminescence and fluorescence of agroinfiltrated leaves infected with *Psy(nptII:lux)*. Luminescence and fluorescence are expressed as relative luminescence units (RLU) and relative fluorescence units (RFU) respectively. All constructs, secreted RFP (secRFP), *Nb*PR3 (PR3), class III Peroxidase (Per25), Pip1 and P69B were co expressed with cytoplasmic GFP and p19. Data points represent biological replicates, derived from the means of three to four samples from the same plant and colours indicate independent experiment. Centre lines are the median, bounds of boxes the 25th and 75th percentiles and whiskers 1.5 x interquartile range (IQR) from 25th and 75th percentiles, outliers are indicated as diamonds. For secRFP-PR3, n=38 from five experiments; For secRFP-Peroxidase and secRFP-Pip1, n=8 from one experiment; For P69B, n=32 from four experiments. Each experiment contained the same number of biological replicates for the negative control (secreted RFP) and proteins of interest; the analysis was thus balanced. Statistical testing was based on one-way blocked ANOVA, using dates of experiments as blocks. Benjamini-Hochberg, false discovery rate-corrected p-values are indicated as: ns>0.05, $*<0.05$, $**<0.01$, $***<0.001$, $****<0.0001$.

4.4 Discussion

In this chapter, we successfully established the agromonasLux disease assay. We showed that agromonasLux was capable of robustly detecting immunity induced by *Sl*FLS3 and

*Nb*PR3 on *Pto* and discussed suitable negative controls. Unfortunately, none of the three tested proteins (Per25, Pip1 and P69B) changed luminescence of agromonasLux and we did thus not identify novel proteins or roles of proteins in apoplastic immunity. AgromonasLux does, however, provide the throughput to test many other proteins for their implication in apoplastic plant immunity. In future, we should study if increased protein secretion and extracellular protein accumulation alone are inducing biotic stress responses. This could indicate a positive feedback loop in apoplastic plant immunity responses and have vital implications for negative controls used in disease assays.

4.4.1 Experimental design

The aim of this section had been to define the experimental design of agromonasLux to reduce bias in data collection and enable reproducibility. We defined the layout of treated plants and used the recently published agromonas assays as a template (Buscaill et al. 2021). As growth facilities limit the number of plants available for experiments, pressure on researchers is high to balance theoretically optimal experimental design with the ability to make significant inferences. Repeating experiments is best practice (Buscaill et al. 2021; Ngou et al. 2021) to ensure robustness of observed effects. Here, I combined blocked ANOVA analysis with experimental replication to collect sufficient evidence to robustly identify effects of small magnitudes. Furthermore, I avoided pseudoreplication (Hurlbert 1984; Blainey et al. 2014) and used samples from different infection events of the same plant only to infer luminescence for each biological replicate. Pseudoreplicates tend to be normally distributed around the mean of an individual rather than the mean of a population (Lazic 2010), meaning that multiple samples from the same individual provide a more accurate assessment of the sampled subject, but do no constitute independent observations of a population. Statistical inference tools such as the student's t-test compare means of populations (Student 1908), in our case young, secRFP- or candidate protein-producing *N.benthamiana* LAB plants. Two hypothetical experiments, one, sampling the same plant 20 times and the other, sampling 20 plants once each, intuitive illustrate that both experiments do not describe their underlying population equally and should not

both be considered as having sample size n=20. This is supported by simulation studies showing that the effective sample size will be overestimated when treating pseudoreplicates as independent (Blainey et al. 2014). Overestimating sample sizes affects the shape of the t-distribution and ultimately statistical inference. More advanced statistical tools such as mixed effect models can account for multiple measurements of individuals in an experiment (Lazar et al. 2016), but were beyond my analytical abilities. Thus, I decided to average repeated measurements of each plant. In future, advanced analytics tools such as mixed effects models could help to take full advantage of generated data by further integrating multiple measurements of the same biological replicate (Lazar et al. 2016).
With the increasing ability in biology to generate large data sets, the need for transparent and robust experimental design and statistical analysis is growing. Open discussions about definitions of replicates and adequate analyses will be paramount to maintain reproducibility. Here, we aimed to set up a transparent experimental design for a robust disease assay to evaluate the effects of agrobacterium-mediated over expression of proteins on bacterial bioluminescence.

4.4.2 Establish AgromonasLux

We established experimental conditions for AgromonasLux as a bioluminescence-based plant disease assay in the *N.benthamiana* - *Pto* pathosystem.
First, we established detection of *Pto(oxb13:lux)*- and *Pto(nptII:lux)*-mediated luminescence in agroinfiltrated *N.benthamiana* leaves. Luminescence was stronger in leaves infected with *Pto(nptII:lux)* compared to *Pto(oxb13:lux)* which was in line with published data of *P.syringae* pv. *phaseolicola* 1302A (*Pph*) transformed with the same Tn7-mini plasmids (Soldan et al. 2020). Luminescence derived from *Pph(oxb13:lux)*, *Pph(nptII:lux)* and *A.tumefaciens(nptII:lux)* is detectable in leaves of french bean (*Phaseolus vulgaris*) and *N.benthamiana* (Soldan et al. 2020; Jutras et al. 2021). Next, we showed that agromonasLux detected effects of immunity proteins *Sl*FLS3 and *Nb*PR3 (Hind et al. 2016, Godson et al., *in preparation*). We had hypothesised that apoplastic immunity effects would be more robustly detected at lower *Pto* infection densities. This hypoth-

esis was inspired by work on *Xanthomonas oryzae* pv. oryzae where effector-mediated enhanced virulence in rice had only been detectable at low infection densities (Han et al. 2011). Along those lines, we reasoned that small effects in plant-pathogen interactions could be masked at high pathogen densities. In our data, however, lower infection densities did not have enough sensitivity to detect effects of apoplastic proteins in plant immunity. On the contrary, we failed to detect differential luminescence upon *Nb*PR3 expression at $OD_{600} = 0.0001$, but succeeded at the higher inoculum $OD_{600} = 0.0010$, despite fewer biological replicates. Furthermore, reduction of luminescence for FLS3 were comparable at $OD_{600} = 0.0001$ and 0.0010 with 3.0-fold and 2.7-fold, respectively.

Next, we compared EV to secreted RFP (secRFP) and found no significant difference in luminescence between EV and secRFP controls (log2FC=0.6, FDR-corrected p-value = 0.21). A reduction in luminescence in agroinfiltrated leaves expressing secRFP was detectable upon *Nb*PR3 expression however. While EV-PR3 had been significant, secRFP-PR3 was not (**Figure 44**). In future, we should aim to study putative small effects of increased protein production in the secretory pathway and apoplastic accumulation of non-immunity proteins such as secRFP to trigger stress in plants that confers increased resistance to bacterial pathogens compared to agroinfiltrated leaves expressing EV. This could provide an explanation as to why we observe immunity-related effects when expressing catalytic mutants of apoplastic hydrolases compared to EV controls when using more biological replicates (Morimoto et al. 2022). While inactive mutants are important to establish necessity of hydrolytic activity of hydrolases in plant immunity, EV controls might not provide a condition comparable to secreted proteins. If increased secretion and extracellular accumulation of proteins induces a stress response, inactive mutants could appear to increase disease resistance relative to EV controls. To test this hypothesis, disease assays such as agromonas or agromonasLux should test if secRFP confers increased disease resistance against *Pto* compared to EV control. If this proves to be the case, considering secRFP over EV as a negative control should increase the robustness of disease assays in the apoplast.

4.4.3 Test tomato candidate proteins in agromonasLux

Finally, we tested candidate proteins from **chapter 2** for their implication in apoplastic immunity against bacteria. Cloning using synthesised gene strings and plasmid backbones designed for transient expression rapidly accelerated the testing of candidate genes in agromonasLux. It took only 19 days to clone and sequence genes and confirm apoplastic accumulation in *N.benthamiana* leaves. Of 15 candidate proteins however, we detected only 7 in AF. Due to time constrains, apoplastic accumulation of candidate proteins had only been evaluated once. Potential reasons for lack of apoplastic accumulation of the remaining eight proteins could be false positively transformed *A.tumefaciens* colonies or a cleaved off C-terminal His tag. Different detection methods such as coomassie stains, protein-specific antibodies or substrate assays could be used in future. Alternatively, gene strings with different tags such as cMyc could be ordered.

After detecting apoplastic accumulation of Per25, Pip1 and P69B, we concluded that their transient expression had no detectable effect on luminescence in *Pto(nptII:lux)*-infected leaf discs of *N.benthamiana* (**Figure 47**). The absence of reduced luminescence due to agrobacterium-mediated protein expression can be interpreted as them not playing a role in apoplastic antibacterial immunity. For Pip1, this had been surprising, since its silencing in transgenic tomato had resulted in increased susceptibility to *Pto* (Ilyas et al. 2015). Possible explanations are insufficient sample size to detect small differences or missing components in *N.benthamiana* necessary for Pip1-mediated increased immunity against *Pto*, unlike in domesticated tomato. In general, a single metric of local luminescence may be insufficient to capture the breadth of possible plant defence responses. Despite a high correlation between luminescence and classic colony counts upon plant defence responses (Fan et al. 2008; Matsumoto et al. 2022), colony counts may show differences that can not be detected by bioluminescence. Furthermore, in the tomato-*Ralstonia solanacearum* pathosystem, bacteria were as abundant in AFs of a resistant cultivar as in a susceptible cultivar but clear difference in disease resistance was detectable in disease symptom severity (Planas-Marquès et al. 2018). In avirulent *P.syringae* pv. *phaseolicola* 1302A (*Pph*), bacteria were found to spread less in french bean leaves compared to virulent *Pph* (Soldan

et al. 2020), indicating that increased immunity could manifest itself in spatial restriction of pathogens rather than reduction of local bacterial density. Bacterial spread in leaves could be evaluated for expressed candidate genes using agromonasLux in future. Especially lignin-forming Per25 is a promising candidate to affect bacterial spread in leaves. If spatial distribution of *Pto(nptII:lux)* and symptom severity are not changed upon expression of candidate proteins, it is possible that the respective protein does not have an effect on immunity against *Pto* in *N.benthamiana*. Candidate genes could, however, play a role in immunity in other pathosystems. Pathogens producing known inhibitors against candidate proteins such as *Phytophthora infestans* which secretes P69 subtilase inhibitors Epi1 and Epi10 could be tested next (Tian et al. 2004; Tian et al. 2005). Despite the facts that P69B releases elicitors from *P.infestans* (Wang et al. 2021a), *P.infestans* secretes P69B inhibitors (Tian et al. 2004; Tian et al. 2005) and P69B being under strong positive selection (**chapter 2**), a role of P69B on *P.infestans* virulence has not been reported. Candidate genes could also be tested for a role in immunity against many other *N.benthamiana* pathogens such as bacterial *Pseudomonas syringae* pv *syringae* or pv. *tabaci*, as well as fungal *Botrytis cinerea* or *Colletotrichum orbiculare*.

5 AlphaFold2 predicts protein complexes at the plant-pathogen interface

5.1 Introduction

Competition between species for limited resources is a well-established concept of a strong driving force in evolutionary biology (Dawkins and Krebs 1979). This competition can turn into antagonistic arms races between species on macro- and micro scales. Macro scales such as speed of movement between prey animals and predators, and micro scales such as interactions between host-derived enzymes and pathogen-derived enzyme inhibitors in the attempt to defend and invade the host, respectively. As early proof of concepts of molecular arms races at the plant-pathogen interface, immunity-related hydrolases have been described to encode measurable traces of molecular evolution, indicative of antagonistic arms races (Bishop et al. 2000; Bishop et al. 2005; Damasceno et al. 2008; Shabab et al. 2008; Kaschani et al. 2010; Kaschani and van der Hoorn 2011). Despite that antagonistic arms races between competing species are a well-established concept in evolutionary biology, surprisingly few apoplastic, immunity-related hydrolases have been studied for traces of arms-race-induced molecular evolution (Godson and van der Hoorn 2021). This had been part of the rationale to study footprints of molecular arms races in **chapter 2** to identify novel immunity-related plant proteins at the plant-pathogen interface in **chapter 4**. A caveat of studying footprints of antagonistic arms races in only one of the involved entities, is that it does not reveal the interaction partner locked in the arms race. This was highlighted in the unsuccessful screen for novel plant proteins in immunity against only one plant pathogen in **chapter 4**. One of the glaring unanswered questions when studying footprints of molecular arms races is which pathogen-derived entity is locked in the arms race, co-evolving with plant-immunity proteins. Interestingly, co-evolution is one central dogma in the groundbreaking protein structure prediction tool AlphaFold2 as well (Jumper et al. 2021). AlphaFold2 was originally designed to predict 3D protein structures of protein monomers and crucial aspects of the structural inference are

based on analysing co-evolution in amino acid multi sequence alignments (MSAs). From amino acid MSAs, AlphaFold2 learned how and which amino acids regularly changed together and inferred that those amino acids are likely to be in close proximity to each other. Shortly after the release of AlphaFold2's trained model to the public, it was detected that co-evolution data within amino acid alignments can be expanded from single proteins to protein complexes (Ko and Lee 2021). Co-evolution of amino acids that are in proximity to each in a protein complex, but are from different chains of the complex, encodes enough information for AlphaFold2 multimer to accurately infer the 3D structures of the overall protein complexes. Soon after, Deepmind implemented protein complex prediction and expanded AlphaFold2 with a module named AlphaFold2 multimer (Evans et al. 2021). Since then, AlphaFold2 has been used to study protein-protein interactions, including a large-scale *in silico* yeast-2-hybrid in *Saccharomyces cerevisiae* (Humphreys et al. 2021). To our knowledge, AlphaFold2 has so far only been applied to study protein complexes from the same species. Molecular co-evolution, however, is not limited to protein complexes from the same species, but is a central characteristic of antagonistic enzyme-inhibitor complexes at the host-pathogen interface. We thus hypothesised that AlphaFold2 multimer is capable of inferring protein complexes from co-evolutionary data encoded in host- and pathogen-derived proteins locked in an antagonistic arms race. In this chapter, we set out to use AlphaFold2 multimer to predict novel enzyme-inhibitor complexes at the apoplastic plant-pathogen interface. A few structural biology metrics generated by AlphaFold2 multimer need to be introduced to be able to follow the analysis of AlphaFold2 multimer outputs. These include the predicted Template Modelling score (pTM), the interface predicted Template Modelling score (ipTM) and the predicted local Distance Difference Test (plDDT). The predicted Template Modelling score is based on the original Template Modelling score which was designed to measure overall fold similarity between two proteins (Zhang and Skolnick 2004). The TM score, as a length-independent metric theoretically ranges from 0 to 1. Scores below 0.17 correspond to a comparison between two unrelated protein folds and scores above 0.5 usually assume similar 3D folds (Zhang and Skolnick 2004). A score of 1 indicates a perfect fit. The TM score

has been used to assess predicted protein structures relative to a ground truth in Critical Assessment of Structure Prediction (CASP) competitions since CASP5 (Zhang and Skolnick 2004). Google Deepmind expanded from the TM score to generate a predicted TM (pTM) score without using structural templates. Therefore, AlphaFold2 compares amino acid positions of its final predicted structure to a distribution of predicted probable structures. From this comparison, AlphaFold2 infers the error of prediction for each residue using a neural network to calculate a pTM value for the whole protein (Jumper et al. 2021). The accuracy of the pTM score compared to TM scores using structural templates was demonstrated in Jumper et al. 2021. Furthermore, AlphaFold2 multimer has generated an interface predicted Template Modelling (ipTM) score that only considers amino acids around the protein-protein interface and predicts their respective TM score (Evans et al. 2021). To score protein complexes, AlphaFold2 multimer combines and balances ipTM and pTM scores by weighing them at factors 0.8 and 0.2 to generate the combined, predicted protein complex score ipTM+pTM (Evans et al. 2021).

The superposition-free local Distance Difference test (lDDT) does not focus on the overall similarity of protein folds (Mariani et al. 2013). Rather than providing a single score for a protein, lDDT scores are calculated for each amino acid separately and range from 0 to 100. To calculate lDDT scores, the predicted structure is aligned to an underlying, resolved structure and preservation of distances to other atoms within a defined radius, the inclusion radius, are calculated for each amino acid. Furthermore, preservation of bond lengths and angle widths relative to the known protein structure are calculated for each amino acid as well. Then, the percentage of preserved distances, lengths and widths is expressed in the lDDT score for each amino acid. Google Deepmind has implemented a predicted lDDT (plDDT) score for modelled protein complexes in AlphaFold2 and has demonstrated strong correlation with calculated lDDT scores when existing structures are available (Jumper et al. 2021). Taken together, AlphaFold2 multimer is capable of producing metrics to judge the accuracy of the predicted overall protein fold (pTM), the accuracy of the fold at the protein-protein interface (ipTM) and the degree to which the local environment of each single amino acid is preserved (plDDT). These scores will be

used to evaluate predicted protein complexes.

In this chapter, we will first provide evidence that AlphaFold2 multimer is capable of predicting high-scoring protein-protein interactions for known enzyme-inhibitor complexes at the plant-pathogen interface. While protein-protein interaction of these enzyme-inhibitor pairs was experimentally confirmed, no structures of their complexes are publicly available. Second we will screen tomato PR proteins against a broad range of pathogen-derived proteins to generate more than 10,000 putative enzyme-inhibitor complexes. Third, we will manually screen a subset of promising protein complexes to identify novel, putative hydrolase inhibitors. Finally, we used biological information such as expression data to identify a selected number of high-quality, novel, putative hydrolase inhibitors at the plant-pathogen interface.

5.2 Materials and Methods:

5.2.1 Protein complex prediction with AlphaFold2 multimer:

Protein complexes were modelled using AlphaFold2 multimer v2.1.1 (Jumper et al. 2021; Evans et al. 2021). The regular AlphaFold2 multimer workflow can be divided into three modules. First a database search and pre-processing module, second, the Evoformer module and third, the structure module (Jumper et al. 2021). To make the overall AlphaFold2 multimer workflow more hardware and time efficient, we modified the pre-processing steps using a custom-made snakemake v6.7.0 workflow (Mölder et al. 2021). When predicting a protein complex using AlphaFold2 multimer, the default workflow first searches for similar template sequences for individual chains in the protein complex (Evans et al. 2021). This template sequence search is identical for AlphaFold2 monomer and AlphaFold2 multimer, except for JackHMMER v3.3 querying the unclustered Uniprot database and additional matching of a hidden markov model, extracted from the Uniref90 MSA, against the AlphaFold2 multimer-specific database PDB seqres (Evans et al. 2021). For our customised workflow, we first predicted monomers for all proteins involved in future protein complex prediction. Then, we manually added the database search against the unclustered Uniprot database with JackHMMer v3.3 for each monomer. Next, we linked all database searches to the output directory of each AlphaFold2 multimer protein complex and ran AlphaFold2 multimer. Then, AlphaFold2 multimer additionally matched HMMs extracted from the Uniref90 MSA against the PDB seqres database for each protein complex. We used the small bfd database and all database were downloaded as instructed in the 'download_all_data.sh' script from AlphaFold2 v2.1.1 release on GitHub. The difference was that we used wget instead of aria2c to download all databases mgnfiy, pdb70, mmCIF, obsolete.dat, pdb_seqres, small bfd, uniclust30, uniprot trembl, uniprot sprot and uniref90.

5.2.2 Analysing output parameters of AlphaFold2 Multimer

For AlphaFold2, accuracy of predicted quality of protein models correlates with the depth and quality of input amino acid multi sequence alignments (MSAs). Both the depth and quality of input amino acid MSA are captured by the mean non-gap amino acid coverage for each protein. The accuracy of predicted quality scores increases with increasing mean non-gap amino acid coverage until it reaches information saturation after a mean non-gap amino acid MSA depth around 100 (Jumper et al. 2021). Mean non-gap amino acid depth for chains of each protein were calculated using the features.pkl output file generated by AlphaFold2 multimer. The analysis was automated in the script af2h_msa_scores.py deposited in the GitHub repository (Homma 2022b). When calculating mean non-gap MSA depths for proteins modelled in several different complexes, we calculated the mean based on its value from all complexes. Deviations within one protein in different complexes were negligible. Analysing the total computing time of AlphaFold2 multimer was done by reading out data stored in the timings.json file of generated protein complexes. To calculate CPU and GPU hours based the timings.json files of modelled monomers, it's necessary to know that we ran all AlphaFold2 monomer computations on the University of Oxford Advanced Research Computing (ARC) facility with eight CPU cores and one GPU at any time. AlphaFold2 multimer computations were run with one CPU core and one GPU at any time.

5.2.3 Selection of tomato and pathogen proteomes and transcriptomes:

For the tomato proteins, we used mature protein sequences from the *S.lycopersicum* ITAG4.0 proteome (Consortium 2012). These were Pip1, P69B, apoplastic enhanced disease susceptibility 1 aspartic protease and class I, III and V chitinases. Used protein sequences of mature tomato proteins in this study can be found in **Table 12**. For tomato pathogens, we selected proteomes and transcriptomes from assemblies GCF_000007805.1 (proteome, transcriptome for *P.syringae* pv. *tomato* DC3000 from pseudomonas.com) (Buell et al. 2003; Winsor et al. 2016); GCF_013112235.1 (proteome and transcriptome of *X.perforans* DSM 18975 from NCBI); GCF_000009125.1 (proteome and transcrip-

Table 12: Sequences of six tomato proteins used in AlphaFold2 multimer modelling

Name	ID	Sequences for modelling (predicted signal peptides and prodomains removed)
Class I Chit.	Solyc09g098540.3.1	DENESFRKQPTLVKTVKGTKMCVKDWECNKLSKFCCNLTITDYLDV DQFELLFTKRNSPVAHAVGFWDYGSFIRAAALYQPLGFGTTGGKKM QMKEIAAFFGHVGSKTSCGYGVATGGPLAYGLCYNKEMSPSQDYCD EYFKLTYPCTPGARYYGRGALPIYWNYNYGAIGEALKLNLLDHPEYI EQNATMAFQAAIWRWMNPMKKGQPSAHDAFVGNWKPTKNDTLSK RVPGFGTTMNILYGDGVCGQGDVDAMNNIVSHYLYYLDLMGVGREE AGPHEVLNCAEQKPFNPTAVAAS
Class III Chit.	Solyc05g050130.4.1	GGIVIYWGQNGNEGTLASTCASNNYAIVNIAFLVVFGNGQTPVLNLA GHCNPSVGGCTRLSNDIRTCQSRGIKVMLSLGGGVGSYSLSSADDARN VANYLWNNYLGGQSTSRPLGDAILDGIDFDIERGTTQHWDELARALS GFSQQRKVYLTAAPQCPFPDSWLNGALSTGLFDYVWVQFYNNPPCQ YSAGSAVNLKSYWNKWTAIQAGKIFLGLPAAPGAAGSGFIPSDVLVS QVLPAINGSPKYGGVMLWSKFYDNGYSSAIKPRV
Class V Chit.	Solyc07g005090.4.1	KGVKGAYYPSWAFSTFPPSSIDTSLFTHIYYAFLVPNNTTFKFDIDDE TSKLLFNFTSTLRSSVKTLFSVGGGGEGPARFSRMASTSVSRLSFIKSS IEVARKYKFDGFDLDWEFPQNKKDMENFAILLNEWRVEVKKESLAT KRPQLLITAAVYFSVDFFLWGEFRSYPVPSINKNLDWINLMFYDYRG SWDTSATGAQAALFDTKSNVSTSYGLSTWIKAGALRSKLIMGLPLYG RTWKLKDPNVDGIGAPAVGVGPGDEGTLTYREIEKFNEENNAKVVY DSATVSAYSVAGTSWIGFDDTNSVAMKLHYAQSQRLRGYFFWAVAG DKDWKISTTAKQSWIIS
P69B	Solyc08g079870.3.1	SLHTTHTPSFLGLQQNMGVWKDSNYGKGVIIGVIDTGIIPDHPSFSDV GMPPPPAKWKGVCESNFTNKCNNKLIGARSYQLGNGSPIDSIGHGTH TASTAAGAFVKGANVYGNADGTAVGVAPLAHIAIYKVCNSVGCSESD VLAAMDSAIDDGVDILSMSLSGGPIPFHRDNIAIGAYSATERGILVSCSA GNSGPSFITAVNTAPWILTVGASTLDRKIKATVKLGNGEEFEGESAYR PKISNATFFTLFDAAKNAKDPSETPYCRRGSLTDPAIRGKIVLCSALG HVANVDKGQAVKDAGGVGMIIINPSQYGVTKSADAHVLPALVVSAA DGTKILAYMNSTSSPVATIAFQGTIIGDKNAPMVAAFSSRGPSRASPGI LKPDIIGPGANILAAWPTSVDDNKNTKSTFNIISGTSMSCPHLSGVAAL LKCTHPDWSPAVIKSAMMTTADTLNLANSPILDERLLPADIYAIGAGH VNPSRANDPGLVYDTPFEDYVPYLCGLKYTDQQVGNLIQRRVNCSE VKSILEAQLNYPSFSIFGLGSTPQTYTRTVTNVGDATSSYKVEVASPE GVAIEVEPSELNFSELNQKLTYQVTFSKTTNSSNPEVIEGFLKWTSNR HSVRSPIAVVSA
Pip1	Solyc02g077040.4.1	VTEVPNSMDWRKRGSVTGVKDQGVCGCCWAFSAAAAIEGAYQIAN NELISLSEQQLLDCSTQNKGCEGGLMTVAYDFLLQNNGGGITTETNY PYEEAQNVCKTEQPAAVTINGYEVVPSDESSLLKAVVNQPISVGIAAN DEFHMYGSGIYDGSCNSRLNHAVTVIGYGTSEEDGTKYWIVKNSWGS DWGEEGYMRIARDVGVDGGHCGIAKVASFPTA
AED1	Solyc08g067100.2.1	LPAQPGIALSTGNYIVTVGIGTPKKDLTLIFDTGSDLTWTQCEPCFKT CFPQQQPIFNPSSSSTYSNISCSSTACSGLKSATGNSPVCSSSTCVYGIQ YGDSSFSIGFFAKDRLTLSATDVFDGFMFGCGQDNKGLFGKTAGLIG LGRDPLSIVSQTSAKFGKYFSYCLPTRRGSNGHLSFGKNGAKSNLQFT PFASSQGTSFYFIDVLGISVGGKSLAISPMVFKNAGTIIDSGTVITRLPS TAYSNLRATFREFMSKYPRAPDLSLLDTCYDLSNYTTISIPKISFNFNG NTKMDIVPNGIFIVNGASQVCLAFAGNGDDDSIGIFGNTQQQTMEIVY DVAGEKLGFGYGGCT

tome of (*R.solanacearum* GMI1000 from NCBI) (Salanoubat et al. 2002); GCF_00014 3535.2 (proteome and transcriptome of *B.cinerea* B05.10 from NCBI) (Van Kan et al. 2017); GCF_000149955.1 (proteome and transcriptome of *F.oxysporum* f. sp. *lycopersici* 4287 from NCBI) (Ma et al. 2010); GCA_020509005.1 (proteome and transcriptome of *C.fulvum* Race5_Kim from NCBI) and GCF_000142945.1 (proteome of *P.infestans* T30-4 from NCBI) (Haas et al. 2009) .

5.2.4 Prediction of small, apoplastic proteins:

To predict small, apoplastic proteins for the seven plant pathogens *P.syringae* pv. *tomato* DC3000, *X.perforans*, *R.solanacearum*, *B.cinerea*, *F.oxysporum* f. sp. *lycopersici*, *C.fulvum* and *P.finfestans*, we downloaded their proteomes as described in **Section 5.2.3**. Next, we used a custom secretion prediction pipeline deposited in out GitHub repository (Homma 2022a). The secretion pipeline produces a summary file, containing information on secretion, apoplastic localisation and transmembrane helices from SignalP5.0, TargetP2.0, ApoplastP1.0.1 and tmhmm.py1.1 for each protein (Sonnhammer et al. 1998; Cock et al. 2009; Sperschneider et al. 2018; Armenteros et al. 2019b; Armenteros et al. 2019a). The pipeline also adds information of molecular weight for each protein using Biopython 1.78 (Cock et al. 2009). We considered proteins to be apoplastic proteins if they were predicted to be secreted by either SignalP5.0 or TargetP2.0 or both and were predicted to be localised in the apoplast by ApoplastP1.0.1. Proteins were considered small if their full length sequence was predicted to <35 kDa. Next, we removed their predicted signal peptides. If a sequences had been predicted by SignalP5.0 to be secreted, we used the mature sequence as predicted by SignalP5.0. If a sequence was only predicted by TargetP2.0 to be secreted, we used the mature sequence as predicted by TargetP2.0. Next we manually added 14 known apoplastic proteins from *C.fulvum* and *F.oxysporum* f. sp. *lycopersici* which did not have identical copies in the predicted proteomes used for this study. These additional 14 proteins included *C.fulvum* proteins AIZ11404.1 (Avr2), AHY02126.1 (Avr5) and AQA29222.1 (Ecp17) and *F.oxypsorum* f. sp. *lycopersici* proteins ALI88770.1 (Six1), UEC48541.1 (partial Six3), BAM37635.1 (Six4), ALI88836.1 (Six6), AIY35187.1 (Six7), ACN69118.1 (Six8), AGG54051.1 (Six10), AGG54052.1 (Six11), ANF89367.1 (Six12), AGG54055.1 (Six14) and APP91304.1 (Six15). Next we filtered all mature, small, putatively apoplastic pathogen-derived proteins against any duplicated amino acid sequences using seqkit (Shen et al. 2016). Removing duplicated amino acid sequences reduced the small, putatively apoplastic protein set to a final list of 1881 pathogen-derived proteins. The jupyter notebook to generate the list of small, apoplastic proteins from output of the secretion prediction

pipeline is deposited in the GitHub repository (Homma 2022b).

5.2.5 Visual scoring of candidate complexes:

Manual scoring of 376 protein complexes was strongly aided by automated processing of AlphaFold2 multimer output with a custom-made python pipeline. The pipeline generated two models: The first model displays both proteins as cartoon and residues coloured based on per residue plDDT values using a reverse rainbow colour scheme. The second model displays host proteins as surface in grey with catalytic residues coloured yellow and pathogen proteins as cartoon and coloured based on per-residue plDDT values using the reverse rainbow colour scheme. Furthermore, disulfide bridges were displayed as sticks and likely hydrogen bonds displayed. The script is deposited in the GitHub repository (Homma 2022b). Next, we manually classified all protein complexes using PyMol into four categories. Category zero contained misfolded, nonsense protein complexes. Category one contained complexes with protein-protein interactions not obstructing the active site. Category two contained complexes with protein-protein interactions obstructing the active site, but without secondary structural support for the putative inhibitory protein segment. Category three contained complexes with protein-protein interactions blocking the active site and having structural support for the putative inhibitory segment such as beta sheets and disulfide bridges. Scoring of protein complexes was independently done by two researchers, which increased the robustness of the assigned categories.

5.2.6 RNAseq data mining, filtering and mapping:

To evaluate gene expression of pathogen-derived proteins during infection, we downloaded publicly available raw-read RNAseq data sets for *P.syringae* pv. *tomato, R.solanacearum, B.cinerea, F.oxysporum* f. sp. *lycopersici* and *C.fulvum*. For *P.infestans*, suitable gene expression data had already been mapped onto the structural annotation used in this study and was available (Huang et al. 2021). For *X.perforans*, we could not identify a suitable *in planta* RNAseq data set.

First, we downloaded sequencing reads for *P.syringae* pv *tomato* infecting *A.thaliana*

leaves (SRR10284966, SRR10284967, SRR10284968, SRR10284969, SRR10284973, SR R10284977 and SRR10284978), *R.solanacearum* infecting tomato petioles (SRR5467166, SRR5467167, SRR5467168), *B.cinerea* infecting tomato leaves (SRR6924534, SRR6924535, SRR6924536), *F.oxysporum* f. sp. *lycopersici* infecting tomato roots (SRR6050413, SRR6050414) and *C.fulvum* infecting tomato leaves (SRR1171035, SRR1171040, SRR 1171043, SRR1171047) using NCBI's sra toolkit fasterq-dump.2.10.9. Second, we labelled each sequencing read by its source of origin with Centrifuge 1.0.4 with NCBI nucleotide non-redundant sequences, last updated 03/03/2018 (Kim et al. 2016). Third, to remove putative host-derived RNA reads, we filtered against tax ids 3700 (Brassicaceae), 3701 (Arabidopsis), 3702 (*A.thaliana*), 4070 (Solanaceae), 4081 (*S.lycopersicum*) and 4107 (Solanum). Afterwards, filtered RNAseq reads were quality trimmed using timmomatic 0.39 ('LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36' for unpaired and paired-end reads) (Bolger et al. 2014b). Last, host-filtered and quality-trimmed reads were mapped onto nucleotide sequences of structural annotations used in this study using Kallisto v0.46.2 (Bray et al. 2016). Due to the heterogeneity of gene expression data in

Table 13: RNAseq data to evaluate gene expression of tomato pathogens during infection

Information	**Bacteria**			**Fungi**			**Oomycetes**
Species	*P.syringae*	*X.perforans*	*R.solanacearum*	*B.cinerea*	*F.oxysporum*	*C.fulvum*	*P.infestans*
Our strain	DC3000	DSM18975	GMI1000	B05.10	4287	Race5_Kim	T30-4
Mapped data[1]	No	No	No	No	No	No	Yes
Data sets[2]	Nobori et al. 2020	-	Khokhani et al. 2017	Müller et al. 2018	Zhao et al. 2018	PRJNA86753	Huang et al. 2020
Data set strains	DC3000	-	GMI1000	Pers. 1794	4287	0WU, IPO1979[3]	88069
Data set Host	*A.thaliana*	-	*S.lycopersicum*	*S.lycopersicum*	*S.lycopersicum*	*S.lycopersicum*	*S.lycopersicum*
Host tissue	leaf	-	petiole	leaf	root	leaf	leaf
Sampling [dpi]	2	-	3	1	1	4, 6, 8, 12,	1, 2, 3
Replicates	3	-	3	3	2	4	4

[1] We reused data if it had previously been mapped to the structural annotation used in this study. [2] In absence of an available publication to cite, we referenced the bioproject found in NCBI's sequence read archive (SRA). [3] Only used for the 6 dpi sample.

host species, infected tissue, time point of sampling, extraction method and replicates, we decided to classify genes as being expressed during infection if they exceed an average gene expression ≥ 2 transcripts per million (tpm). For reference, EBI's gene expression atlas considers the lower cut off as 0.5 tpm.

5.2.7 Clustering of proteins of 37 plant pathogens in Orthogroups

To compare homologs of novel putative hydrolase inhibitors, we clustered whole proteomes of 37 scientifically and agriculturally important plant pathogens (**Table 14**) (Dean et al. 2012; Mansfield et al. 2012; Kamoun et al. 2015). Proteins were clustered in orthogroups based on sequence similarity using Orthofinder2.0 (Emms and Kelly 2019). We used Orthofinder2.0 to infer the underlying species tree from whole proteomes. In orthofinder, we used dendroblast and fasttree 2 to infer approximately-maximum likelihood phylogenetic trees (Kelly and Maini 2013; Price et al. 2010). The species tree was inferred by the Species Tree Inference from All Genes (STAG) method using closely-related single- and multi-copy orthogroups and the species tree's root was inferred from gene duplication events using the Species Tree Root Inference from Gene Duplication Events (STRIDE) method (Emms and Kelly 2017; Emms and Kelly 2018). The entire workflow to infer orthogroups and the rooted species tree is encoded in the Orthofinder2.0 command line tool. The phylogenetic tree and distribution of orthogroups between different species were visualised in iTOL v5 using the annotation template (Letunic and Bork 2021).

Table 14: Used proteomes of 37 plant pathogens

Type	Species	Strain	Annotation	Reason
Fungus	*Magnaporthe oryzae*	70-15	GCF_000002495.2_MG8	(Dean et al. 2012)
Fungus	*Botrytis cinerea*	B05.10	GCF_000143535.2_ASM14353v4	(Dean et al. 2012)
Fungus	*Puccinia graminis*	CRL 75-36-700-3	GCF_000149925.1_ASM14992v1	(Dean et al. 2012)
Fungus	*Fusarium graminearum*	NRRL 31084	GCF_000240135.3_ASM24013v3	(Dean et al. 2012)
Fungus	*Fusarium oxysporum*[1]	4287	GCF_000149955.1_ASM14995v2	(Dean et al. 2012)
Fungus	*Blumeria graminis*[2]	-	GCA_900519115.1_Bgt_genome_v3.16	(Dean et al. 2012)
Fungus	*Zymoseptoria tritici*	IPO323	GCF_000219625.1_MYCGR_v2.0	
Fungus	*Colletotrichum graminicola*	M1.001	GCF_000149035.1_C_graminicola_M1_001_V1	(Dean et al. 2012)
Fungus	*Ustilago maydis*	521	GCF_000328475.2_Umaydis521_2.0	(Dean et al. 2012)
Fungus	*Cercospora beticola*	09-40	GCF_002742065.1_CB0940_V2	
Fungus	*Cladosporium fulvum*	Race5_Kim	GCA_020509005.1_Cfulv_R5_v4	
Oomycete	*Phytophthora infestans*	T30-4	GCF_000142945.1_ASM14294v1	(Kamoun et al. 2015)
Oomycete	*Phytophthora ramorum*	Pr-102	GCA_020800215.1_PR-102_v3_p	(Kamoun et al. 2015)
Oomycete	*Phytophthora sojae*	P6497	GCF_000149755.1_P.sojae_V3.0	(Kamoun et al. 2015)
Oomycete	*Phytophthora capsici*	LT1534-B	GCA_016618375.1_Pcap_4.1	(Kamoun et al. 2015)
Oomycete	*Phytophthora cinnamomi*	GKB4	GCA_018691715.1_ASM1869171v1	(Kamoun et al. 2015)
Oomycete	*Phytophthora parasitica*	INRA-310	GCF_000247585.1_PP_INRA-310_V2	(Kamoun et al. 2015)
Oomycete	*Pythium insidiosum*	Pi-s	GCA_001029375.1_Pythium_insidiosum_1.0	(Kamoun et al. 2015)
Oomycete	*Albugo candida*	Ac Nc2	GCA_001078535.1_ASM107853v1	(Kamoun et al. 2015)
Bacterium	*Ralstonia solanacearum*	GMI1000	GCF_000009125.1_ASM912v1	(Mansfield et al. 2012)
Bacterium	*Ralstonia solanacearum*	IBSBF1503	$GCF_0$01587155.1_ASM158715v1	(Mansfield et al. 2012)
Bacterium	*Agrobacterium tumefaciens*	12D1	GCF_003667905.1_ASM366790v1	(Mansfield et al. 2012)
Bacterium	*Xanthomonas oryzae*[3]	ICMP3125	GCF_004136375.1_ASM413637v1	(Mansfield et al. 2012)
Bacterium	*Xanthomonas campestris*[4]	85-10	GCF_000009165.1_ASM916v1	(Mansfield et al. 2012)
Bacterium	*Xanthomonas campestris*[5]	ATCC 33913	GCF_000007145.1_ASM714v1	(Mansfield et al. 2012)
Bacterium	*Xanthomonas axonopodis*[6]	CHN01	GCF_020866885.1_ASM2086688v1	(Mansfield et al. 2012)
Bacterium	*Xanthomonas perforans*	DSM 18975	GCF_013112235.1_ASM1311223v1	
Bacterium	*Erwinia amylovora*	CFPB1430	GCF_000091565.1_ASM9156v1	(Mansfield et al. 2012)
Bacterium	*Xylella fastidiosa*	Temecula1	GCF_000007245.1_ASM724v1	(Mansfield et al. 2012)
Bacterium	*Dickeya dadantii*	3937	GCF_000147055.1_ASM14705v1	(Mansfield et al. 2012)
Bacterium	*Dickeya solani*	PPO 9019	GCF_002846995.1_ASM284699v1	
Bacterium	*Pectobacterium carotovorum*	WPP14	GCF_013488025.1_ASM1348802v1	(Mansfield et al. 2012)
Bacterium	*Pectobacterium atrosepticum*	21A	GCF_000740965.1_ASM74096v1	(Mansfield et al. 2012)
Bacterium	*Pseudomonas syringae* tom.[7]	DC3000	GCF_000007805.1	Mansfield et al. 2012
Bacterium	*Pseudomonas syringae* syr.[8]	B728A	GCF_000012245.1_ASM1224v1	(Mansfield et al. 2012)
Bacterium	*Pseudomonas syringae* tab.[9]	ATCC 11528	GCF_000145945.2_ASM14594v2	(Mansfield et al. 2012)
Bacterium	*Pseudomonas syringae* pha.[1]0	1448A	GCF_000012205.1_ASM1220v1	(Mansfield et al. 2012)

[1] *F.oxysporum* f. sp. *lycopersici*. [2] *B.graminis* f. sp. *tritici*. [3] *X.oryzae* p. *oryzae*. [4] *X.campestris* pv. *vesicatoria*. [5] *X.campestris* pv. *campestris*. [6] *X.axonopodis* pv. *manihotis*. [7] *P.syringae* pv. *tomato*. [8] *P.syringae* pv. *syringae*. [9] *P.syringae* pv. *tabaci*. [1]0 *P.syringae* pv. *phaseolicola*.

5.3 Results

5.3.1 AlphaFold2 models experimentally validated protein complexes from the apoplastic plant-pathogen interface with high predicted quality

To establish AlphaFold2 multimer's ability to predict protein complexes locked in inter-species, antagonistic arms races, we first modelled well-studied protein complexes from the apoplastic plant-pathogen interface. For this first experiment, we modelled two well-studied inhibitor-protease complexes: tomato papain-like cysteine protease *Sl*Pip1 with *Phytophthora infestans*' cognate cystatin-like inhibitor *Pi*EpiC2B and tomato subtilase-like *Sl*P69B and *P.infestans*' kazal-like *Pi*Epi1a (Tian et al. 2004; Tian and Kamoun 2005; Tian et al. 2007). AlphaFold2 multimer predicted the complexes of the positive controls *Sl*Pip1-*Pi*EpiC2B and *Sl*P69B-*Pi*Epi1a with high ipTM+pTM scores (highest-scoring models 0.92 and 0.93 respectively) (**Figure 48A**). The interfaces of reverse, incompatible complexes *Sl*Pip1-*Pi*Epi1a and *Sl*P69B-*Pi*EpiC2B were scored with low confidence (0.48

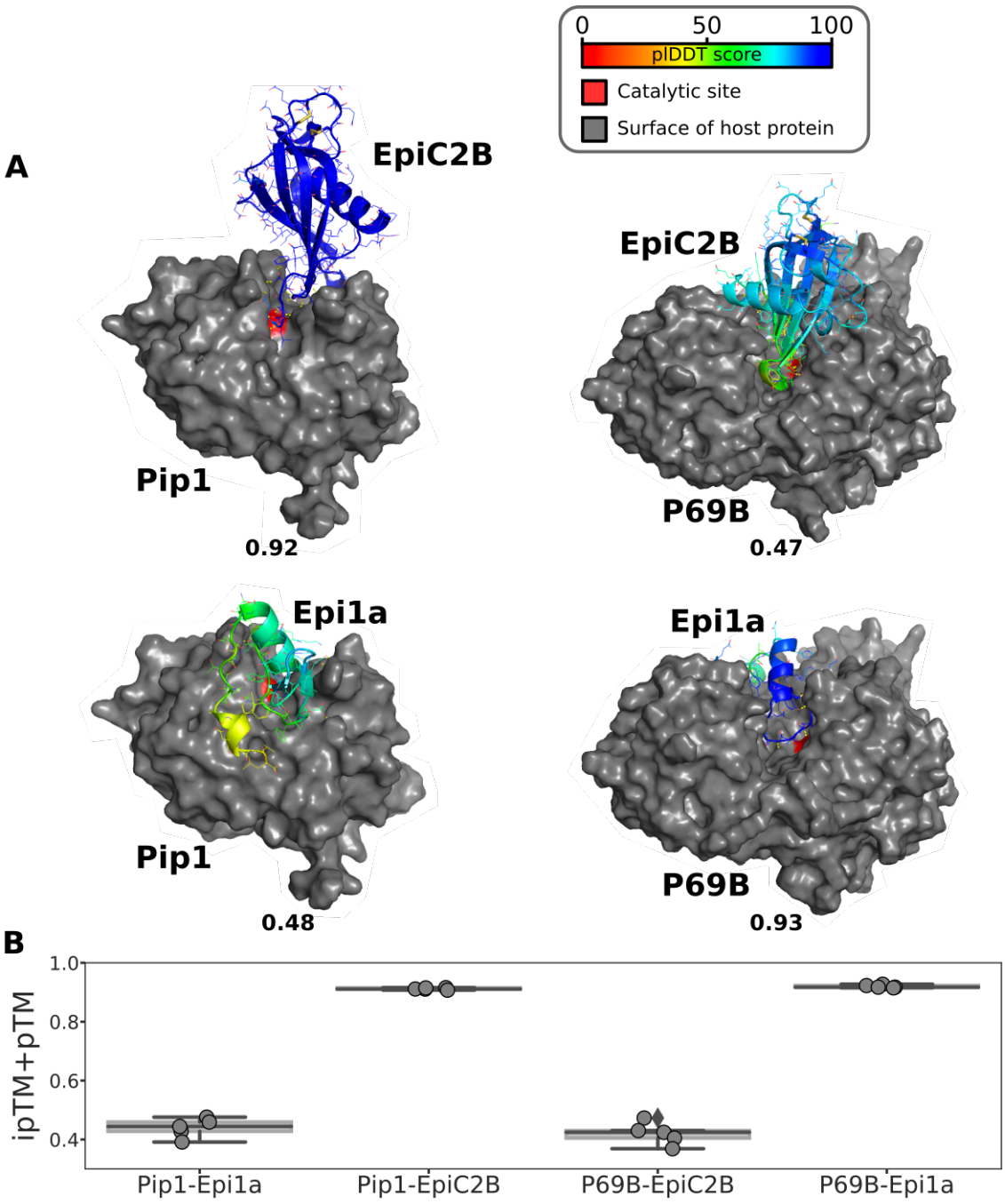


Figure 48: AlphaFold2 predicted quality scores reflect validated inhibitor-protease interactions. (**A**) Predicted protein complexes of compatible (*Sl*Pip1-*Pi*EpiC2B, *Sl*P69B-*Pi*Epi1a) and incompatible (*Sl*Pip1-*Pi*Epi1a, *Sl*P69B-*Pi*EpiC2B) inhibitor-enzyme interactions. Values under complexes reflect the ipTM+pTM scores of the best predicted model for each complex with the scores ranging from 0 (worst) to 1 (best). Plant proteins *Sl*Pip1 and *Sl*P69B are coloured grey and their catalytically active residues red. Amino acids of *Phytophthora infestans*-derived *Pi*EpiC2B and *Pi*Epi1a are rainbow-coloured based on their respective plDDT scores ranging from 0 (worst) to 100 (best). Relative sizes between different protein complexes are not proportionate. (**B**) Predicted quality scores ipTM+pTM of all five models for each modelled inhibitor-protease complex. Centre lines are the median, bounds of boxes the 25th and 75th percentiles and whiskers 1.5 x interquartile range (IQR) from 25th and 75th percentiles. Diamonds represent outliers.

and 0.47 for their highest-scoring model respectively) (**Figure 48A**). For each of the four complexes, all five models were consistently assigned similar ipTM+pTM scores (**Figure 48B**).

We next evaluated the generated models. Cystatins like *Pi*EpiC2B commonly form a tripartite wedge inhibitory structure consisting of the N-terminus with a conserved glycine and two loops (L1 and L2) (Stubbs et al. 1990; Alvarez-Fernandez et al. 2005; Chu et al. 2011), with a conserved QxVxG motif in L1 and the conserved aromatic residue tryptophane or histidine in L2. In the positive control *Sl*Pip1-*Pi*EpiC2B, the tripartite wedge was located in the active site and formed 13 hydrogen bonds (H-bonds) with *Sl*Pip1. In the negative control *Sl*P69B-*Pi*EpiC2B, the tripartite wedge assumed a similar fold as in the *Sl*Pip1-*Pi*EpiC2B complex and clearly blocked the active site, but only eight H-bonds were formed. No collision of amino acids between *Sl*P69B and *Pi*EpiC2B are detectable. The low predicted ipTM+pTM and plDDT scores at the interface however indicate high uncertainty of the predicted complex. Kazal inhibitors like *Pi*Epi1 make 12 contacts in the active site of subtilases (Read et al. 1983; Laskowski et al. 1987; Lu et al. 1997; Lu et al. 2001). In the positive control *Sl*P69B-*Pi*Epi1a, the reactive site loop (RSL) in *Pi*Epia1 forms eleven H-bonds in the active site and the predicted P1 residue (aspartic acid) forms six of them (Tian et al. 2005). It is typical for P1 residues of kazal inhibitors to form half of the hydrogen bonds in the active site (Derache et al. 2012). In addition to the 12 contacts, we identified a glycine forming an H-bond with the hypervariable site III histidine of *Sl*P69B described in **chapter 2**. In the negative control *Sl*Pip1-*Pi*Epi1a, the twelve kazal contact positions again formed the RSL, which was located at the edge of *Sl*Pip1's active site. The P1 aspartic acid was not located next to the catalytic cysteine and formed only one of the six H-bonds. Furthermore, P3 tyrosine sterically collided with two *Sl*Pip1 residues and the additional glycine is located away from the active site. While AlphaFold2 multimer modelled the inhibitors near or inside the active site in all four complexes, the strong contrast between positive and negative controls are the predicted superposition scores (ipTM+pTM) and predicted local accuracy (plDDT) scores. We thus concluded that for these two positive and negative controls, predicted quality

scores are a fast and accurate metric to distinguish between biologically true and false inhibitor-enzyme complexes.

To test AlphaFold's ability as a screening tool, we conducted a larger *in silico* experiment by modelling 28 experimentally-validated protein complexes at the apoplastic plant pathogen interface (overview in **Table 15**). To understand the accuracy of predicted

Table 15: Experimental evidence for direct interaction of apoplastic proteins at the plant-pathogen interface

Plant		**Pathogen**		
Protein	Species	Protein	Species	Evidence
*Gm*EGaseA	*G.max*	*Ps*Gip1	*P.sojae*	NGE[1] Ham et al. 1997
*At*EGase	*A.thaliana*	*Hg*30C02	*H.glycines*	Y2H[2] Hamamouch et al. 2012, BiFC[3] Hamamouch et al. 2012
*Gm*Gip1	*G.max*	*Ps*XEG1	*P.sojae*	co-IP[4] Ma et al. 2017
*Gm*Gip1	*G.max*	*Ps*XLP1	*P.sojae*	co-IP Ma et al. 2017
*Nb*Gip2	*N.benthamiana*	*Pp*XEG1	*P.parasitica*	co-IP Ma et al. 2017
*Nb*Gip2	*N.benthamiana*	*Pp*XLP1	*P.parasitica*	co-IP Ma et al. 2017
*Spim*Rcr3	*S.pimpinellifolium*	*Gr*VAP1	*G.rostochiensis*	co-IP Lozano-Torres et al. 2012, cABPP[5] Lozano-Torres et al. 2012
*Spim*Rcr3	*S.pimpinellifolium*	*Cf*Avr2	*C.fulvum*	co-IP Rooney et al. 2005, cABPP Rooney et al. 2005, cABPP MS van Esse et al. 2008
*Spim*Rcr3	*S.pimpinellifolium*	*Ps*Cip1	*P.syringae*[6]	cABPP Shindo et al. 2016
*Spim*Rcr3	*S.pimpinellifolium*	*Pi*EpiC2B	*P.infestans*	co-IP Song et al. 2009, cABPP Song et al. 2009
*Sl*Pip1	*S.lycopersicum*	*Cf*Avr2	*C.fulvum*	cABPP van Esse et al. 2008
*Sl*Pip1	*S.lycopersicum*	*Pi*EpiC2B	*P.infestans*	co-IP Tian et al. 2007
*Sl*Pip1	*S.lycopersicum*	*Pi*EpiC1	*P.infestans*	co-IP Tian et al. 2007
*Sl*Pip1	*S.lycopersicum*	*Ps*Cip1	*P.syringae*[6]	cABPP Shindo et al. 2016
*Sl*C14	*S.lycopersicum*	*Cf*Avr2	*C.fulvum*	cABPP van Esse et al. 2008
*Sl*C14	*S.lycopersicum*	*Pi*EpiC1	*P.infestans*	cABPP Shindo et al. 2016
*Sl*C14	*S.lycopersicum*	*Pi*EpiC2B	*P.infestans*	cABPP Shindo et al. 2016
*Sl*C14	*S.lycopersicum*	*Ps*Cip1	*P.syringae*[6]	cABPP Shindo et al. 2016
*Zm*CP2	*Z.mays*	*Um*Pit2	*U.maydis*	Y2H Mueller et al. 2013, co-IP Mueller et al. 2013
*Zm*CP1A	*Z.mays*	*Um*Pit2	*U.maydis*	Y2H Mueller et al. 2013
*Zm*CP1B	*Z.mays*	*Um*Pit2	*U.maydis*	Y2H Mueller et al. 2013
*Zm*XCP2	*Z.mays*	*Um*Pit2	*U.maydis*	Y2H Mueller et al. 2013
*Cs*RD21a	*C.sinensis*	*Cl*SDE1	*C.Liberibacter*[7]	Y2H Clark et al. 2018, co-IP Clark et al. 2018, cABPP Clark et al. 2018
*Sl*P69B	*S.lycopersicum*	*Pi*Epi1	*P.infestans*	co-IP Tian et al. 2004
*Sl*P69B	*S.lycopersicum*	*Pi*Epi1a	*P.infestans*	co-IP Tian and Kamoun 2005
*Sl*P69B	*S.lycopersicum*	*Pi*Epi10	*P.infestans*	co-IP Tian et al. 2005
*Zm*POX12	*Z.mays*	*Um*Pep1	*U.maydis*	BiFC (Hemetsberger et al. 2012), Y2H (Hemetsberger et al. 2012)

[1] NGE: native gel electrophoresis of purified proteins. [2] Y2H: yeast-to-hybrid. [3] BiFC: Bimolecular fluorescence complementation. [4] Co-IP: Co-immunoprecipitation. [5] cABPP: competitive ABPP. [6] *P.syringae* pv. *tomato*. [7] *C.liberibacter asiaticus*.

quality scores for compatible and incompatible enzyme-inhibitor complexes, we investigated the mean non-gap amino acid coverage for all host- and pathogen derived proteins. For all host proteins, mean non-gap MSA depth was well above saturation of information (**Figure 49A**). For pathogen-derived proteins however, only 13 of the 18 modelled

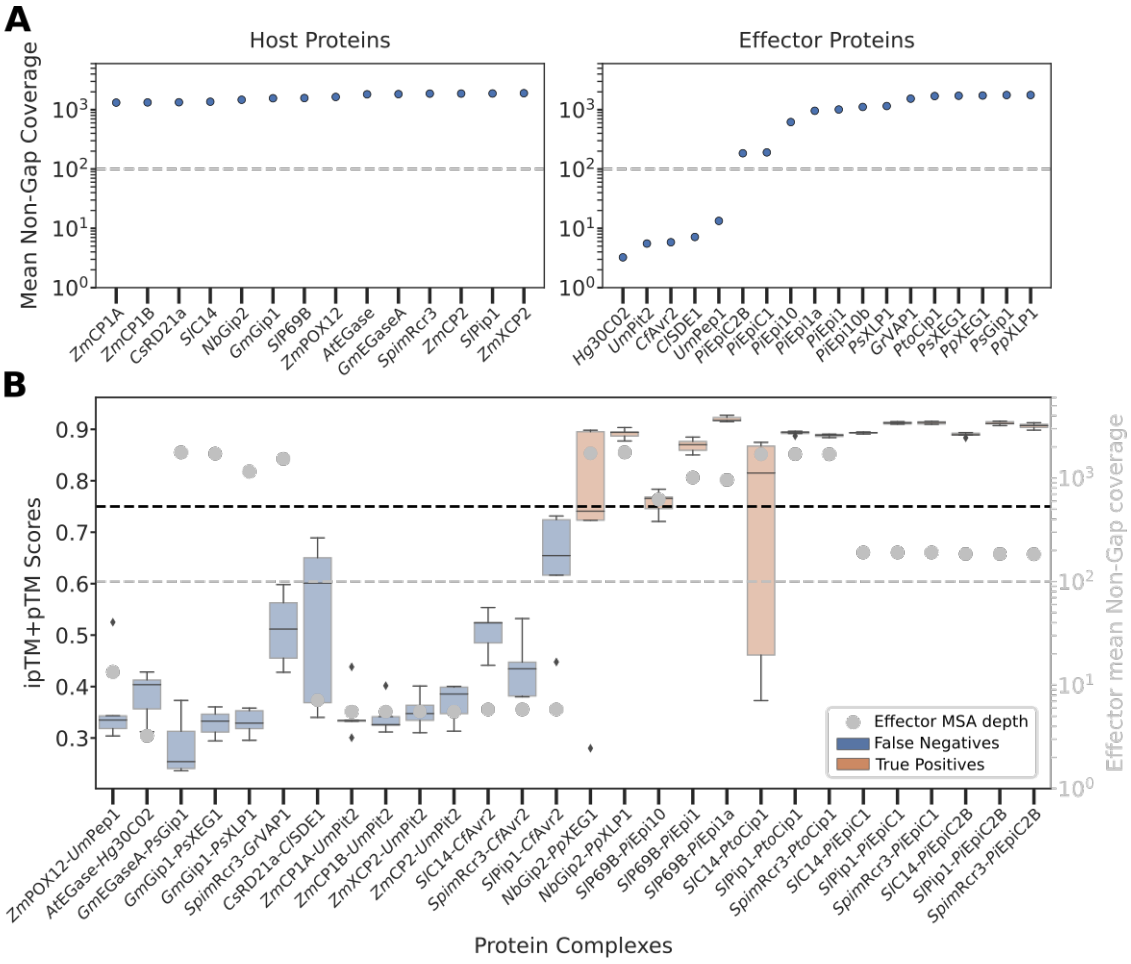


Figure 49: AlphaFold2 models experimentally validated protein complexes at the apoplastic plant-pathogen interface with high predicted quality. (A, B) The grey dashed horizontal line indicates a mean non-gap MSA depth of 100 and the black horizontal dashed line indicates an ipTM+pTM score of 0.75. **(A)** Mean non-gap amino acid coverage in protein MSAs used to predict protein complexes between host target and pathogen effector proteins. **(B)** The ipTM+pTM scores of five models for a selection of experimentally validated protein complexes at the apoplastic plant-pathogen interface are displayed. Data points in grey highlight the mean non-gap amino acid coverage in the input MSA for the pathogen-derived effector. Centre lines are the median, bounds of boxes the 25th and 75th percentiles and whiskers 1.5 x interquartile range (IQR) from 25th and 75th percentiles. Diamonds represent outliers.

proteins showed a mean non-gap MSA depth above the information saturation of 100. The five remaining pathogen-derived, apoplastic effectors *Hg*30C02, *Um*Pit2, *Cf*Avr2, *Cl*SDE1 and *Um*Pep1 showed a mean non-gap MSA depth of 3.2-13.3 (**Figure 49A**). Since the shallow MSA depth of these effectors could be a bottleneck in accurately predicting protein complex scores, we next analysed ipTM+pTM predicted quality scores for our 28 positive controls and compared them to mean non-gap amino acid MSA depth of

the effectors. 14 of the 28 tested positive controls contained at least one complex with a predicted score higher than 0.75 (**Figure 49B**), of which 11 consistently scored higher than 0.75, while three scored both above and below 0.75. For instance, *Sl*C14-*Pto*Cip1 complexes received ipTM+pTM scores ranging from 0.37 to 0.87. All these 14 protein complexes had an effector mean non-gap MSA depth above information saturation. The predicted quality scores for complexes between *Sl*C14 and *Pi*EpiC2B and *Pi*EpiC1, as well as *Spim*Rcr3 and *Pi*EpiC2B and *Pi*EpiC1 did not identify *Pi*EpiC2B as a stronger inhibitor of the two PLCPs (**Figure 50**), in contrast to experimental findings (Hörger & van der Hoorn, *unpublished*). Moreover, ipTM+pTM scores did not identify *PiEpiC2B*

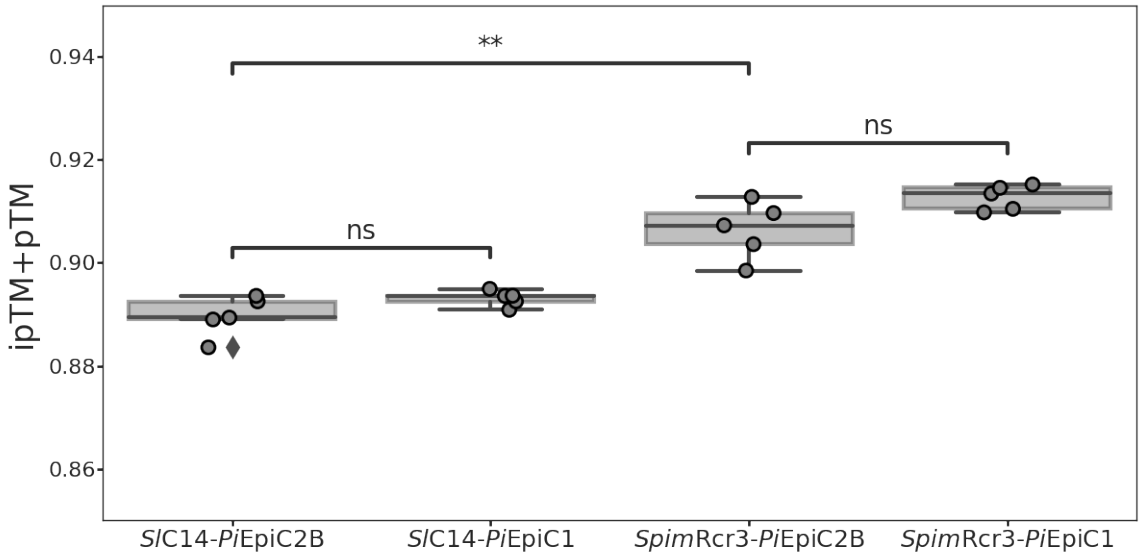


Figure 50: Predicted quality scores of inhibitor-protease complexes by AlphaFold2 do not reflect the strength of an inhibition. Data points reflect the ipTM+pTM scores of the five differnt models for each inhibitor-protease complex. Significance of comparisons was based on Benjamini-Hochberg-based multiple testing-corrected p-values from two-sided, unpaired t-tests. Centre lines are the median, bounds of boxes the 25th and 75th percentiles and whiskers 1.5xinterquartile range (IQR) from 25th and 75th percentiles.

to be a stronger inhibitor of *Sl*C14 than *Spim*Rcr3, in contrast to experimental evidence (Kaschani et al. 2010). The contrary, ipTM+pTM scores for *Sl*C14-*Pi*EpiC2B were lower than for *Spim*Rcr3-*Pi*EpiC2B (**Figure 50**). Concluding, AlphaFold2 multimer predicted quality scores did not reflect the strength of inhibition. The remaining 14 interactions have ipTM+pTM scores consistently below 0.75. Considering that all 14 protein complexes are supported by experimental evidence, we consider them false negatives, missed

by AlphaFold2 multimer. Importantly, ten of these 14 false negatives have an effector mean non-gap MSA depth below 15. For the remaining four false negatives, however, there is sufficient sequence information, yet the predicted quality scores are low. Three of these false negatives are glucanases with cognate inhibitors and include soybean β-1,3-endoglucanase A (*Gm*EgaseA) and its cognate *P.sojae*-derived glucanase inhibitor protein 1 (*Ps*Gip1), soybean glucanase inhibitor protein 1 (*Gm*Gip1) and both its targets, *P.sojae* xyloglucan-specific endoglucanase 1 (*Ps*XEG1) and xyloglucan-specific endoglucanase-like protein 1 (*Ps*XLP1) (**Figure 49B**). The fourth false negative is tomato *Spim*Rcr3 and a nematode-derived inhibitor *Gr*VAP1, an interaction our lab was unable to confirm (Kourelis et al. 2020). To evaluate the false positive rate, we tested all interactions between the 18 inhibitors and 14 enzymes of the 28 known complexes, resulting in 224 additional complexes. These 224 complexes contained 16 likely inhibitor-enzyme complexes with a mean ipTM+pTM score above 0.75, including 14 PLCP-inhibitor complexes, *Nb*Gip2-*Ps*XEG1 and *Sl*P69B-*Pi*Epi10b (**Figure 51**). None of these 16 complexes have been

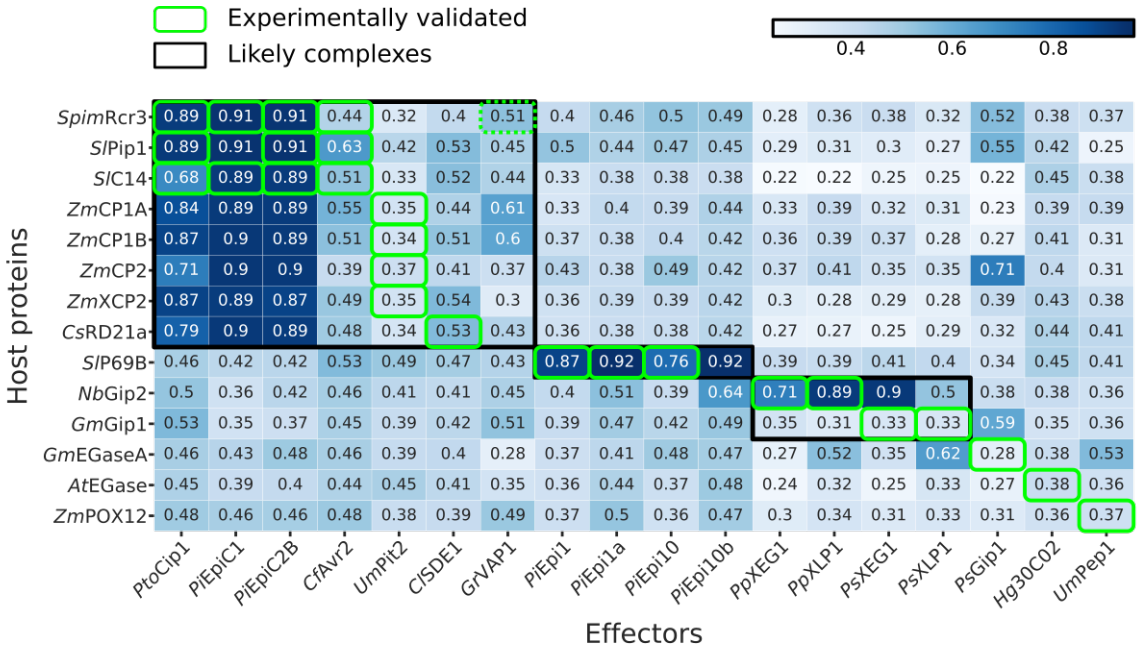


Figure 51: AlphaFold2 models experimentally validated protein complexes at the apoplastic plant-pathogen interface with high predicted accuracy. A heatmap showing the mean ipTM+pTM scores of all five models predicted by AlphaFold2 multimer for 252 protein complexes. Host and effector proteins were selected based on experimentally-validated inhibitor-enzyme complexes at the apoplastic plant-pathogen interface (**Table 15**).

validated experimentally, but they belong to classes of enzymes and inhibitors known to interact. *Pi*Epi10b for instance, is the kazal-like domain of *Pi*Epi10 with the highest predicted affinity to inhibit subtilases (Tian et al. 2005). Aside from the 28 positive controls and 16 high-scoring predicted complexes of known antagonistic enzyme and inhibitor classes, all 208 remaining complexes were assigned ipTM+pTM scores below 0.75 (**Figure 51**), consistent with being incompatible complexes. We thus set the ipTM+pTM threshold for a modelled protein-protein pair to be considered as a putative inhibitor-enzyme complex at 0.75.

Taken together, we conclude that AlphaFold2 multimer can reliably identify experimentally-validated protein complexes at the apoplastic plant-pathogen interface when supplied with mean non-gap protein MSAs of at least 100 (recall of 77.78%). Given sufficient sequence information, AlphaFold2 is especially reliable (recall of 93.33%) at predicting protein complexes between apoplastic proteases such as papain-like cysteine proteases (PLCPs), subtilases and cognate inhibitors with high predicted quality scores (**Figure 49B**). An additional 16 inhibitor-enzyme complexes were identified between protein classes known to form complexes. None of the other, incompatible interactions had ipTM+pTM scores >0.75. An incompatible interaction, and thus a potential false positive, was defined by a pair of two classes of proteins without known protein-protein interactions. The false positive rate in our data set was thus 0.0%, since no predicted protein complexes for incompatible interactions received ipTM+pTM scores above 0.75, For our limited data set of compatible and incompatible inhibitor-enzyme pairs at the plant pathogen interface, AlphaFold2 multimer quality score-based identification of inhibitor-enzyme complexes showed a high rate of true positives and a low rate of false positives. We thus attempted next to identify novel, putative apoplastic pathogen-derived proteins targeting plant immunity-related hydrolases in the apoplast, based on AlphaFold2 multimer-derived predicted protein complex quality scores.

5.3.2 Modelling over 11,000 protein pairs with AlphaFold2 multimer

To discover putative novel apoplastic interactions at the plant pathogen interface, we analysed predicted secretomes of seven tomato plant pathogens. The pathogens were selected based on scientific and agricultural importance in addition to covering a broad range of different species. We chose the bacterial tomato pathogens *Pseudomonas syringae* pv. *tomato*, *Ralstonia solanacearum* and *Xanthomonas perforans*, the fungal tomato pathogens *Botrytis cinerea*, *Fusarium oxysporum* f. sp. *lycopersici* (*Fol*) and *Cladosporium fulvum* and the oomycete *Phytophthora infestans*. *P.syringae*, *R.solanacearum*, *B.cinerea*, *F.oxysporum* and *P.infestans* are amongst the most important plant pathogens in molecular plant pathology (Dean et al. 2012; Mansfield et al. 2012; Kamoun et al. 2015). For all seven pathogens, we predicted proteins that are small (<35 kDa) and likely to be secreted and remaining in the apoplast. Secretion or apoplastic localisation was predicted by at least one of signalP5.0, TargetP2.0 or ApoplastP1.0.1 (Sperschneider et al. 2018; Armenteros et al. 2019b; Armenteros et al. 2019a). The predicted apoplastic secretomes contained a total of 1,881 proteins and were 3-5-fold larger for eukaryotic pathogens than for bacterial pathogens (**Figure 52**). This selection will not include all possible

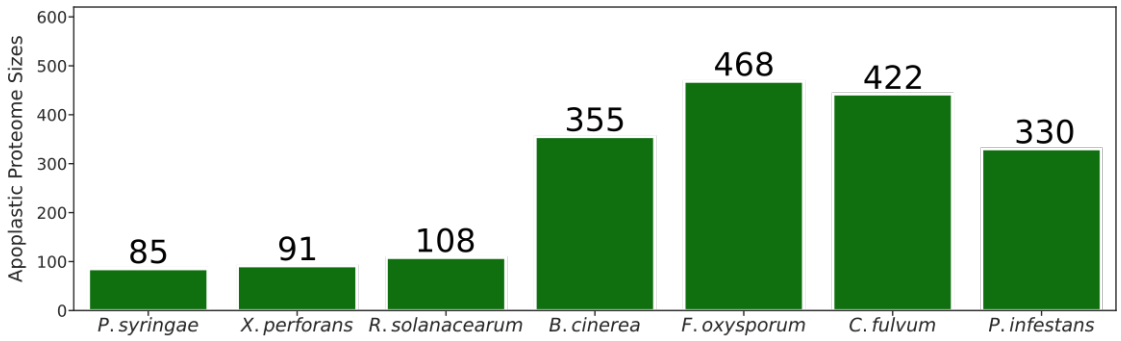


Figure 52: Prediction of small apoplastic secretomes of seven tomato pathogens. Apoplastic localisation was based on our secretion prediction pipeline.

pathogen-derived apoplastic inhibitors, but comprised a heuristic selection that allows for a computationally feasible *in silico* screen against six tomato proteins.

We restricted our analysis to apoplastic plant hydrolases implicated in plant immunity and with a clear substrate binding groove that aids detection of enzyme inhibitors. Based

on these criteria, we selected the immune protease Pip1 (Solyc02g077040.4.1) as well as four bona fide PR proteins: a class I chitinase (PR3 - Solyc09g098540.3.1), a class III chitinase (PR8 - Solyc05g050130.4.1), a class V chitinase (PR3 - Solyc07g005090.4.1) and the subtilase P69B (PR7 - Solyc08g079870.3.1) (van Loon et al. 2006; Sinha et al. 2014). P69B (PR7) is one of the most abundant proteins in Bion-induced AFs of tomato, is implicated in immunity by activating the tomato immunity protease Rcr3 and releasing an elicitor peptide from the *P.infestans* effector PC2 (Paulus et al. 2020; Wang et al. 2021a). Despite molecular evidence of an antagonistic arms race, only the two *P.infestans* kazal-like *Pi*Epi1 and *Pi*Epi10 are proteinaceous inhibitors of P69B (Tian et al. 2004; Tian et al. 2005). The class III chitinase (PR8) was the most abundant chitinase in *Solanum peruvianum*, and probably has dual activity as a chitinase and lysozyme (Terwisscha van Scheltinga et al. 1995). No proteinaceous, pathogen-derived inhibitor is known for PR8. Class I and V chitinases are PR3 proteins and the class V chitinase (Solyc07g005090.4.1) was only detected in AF from Bion-induced *Solanum pennellii*. Class I chitinases are under positive selection, indicative of antagonistic arms races with pathogen-derived inhibitors (Bishop et al. 2000). Chitinases and their ability to degrade chitin as part of the fungal cell wall highlight them as targets for pathogen-derived interference during attempted host colonisation. No proteinaceous chitinase inhibitors are known, however and different pathogen defence strategies against chitinases such as sequestering chitin elicitors or direct degradation of chitinases have been described (reviewed in (Buscaill and van der Hoorn 2021)). Pip1 was the most abundant PLCP in apoplastic fluid extracts of *S.peruvianum* and plays a role in the defence of *S.lycopersicum* against *P.syringae* pv. *tomato, C.fulvum* and *P.infestans* (Ilyas et al. 2015). Last, we added pepsin-like apoplastic enhanced disease susceptibility 1 (AED1) as a negative control as it is implicated in suppression of systemic acquired resistance (SAR) and was thus not expected to be targeted for inhibition by plant pathogens (Breitenbach et al. 2014).

During the modelling of all putative enzyme-inhibitor pairs, AlphaFold2 multimer repeatedly produced different errors when attempting to model XP001552834.1 and XP018232955.1 in combination with any of the six tomato proteins. We decided to not pursue mod-

elling protein complexes with these two pathogen-derived proteins. The remaining 1,879 pathogen-derived proteins were tested in combination with each of the six host proteins (11,274 protein pairs). Overall, modelling 11,274 protein pairs with our custom-made AlphaFold2 workflow required 13,244.20 CPU hours (1.51 CPU years) and 8,117.84 GPU hours (0.93 GPU years). That equals 1.17 CPU hours and 0.72 GPU hours per protein complex. These immense hardware requirements were made feasible by using the University of Oxford Advanced Research Computing (ARC) facility.

We next investigated the protein MSA depth of all proteins involved in the screen. All six tomato proteins, had a mean non-gap MSA depth far above information saturation, ranging from 1,338.8 to 1,872.8 (**Figure 53A**). Thus, protein sequence information for

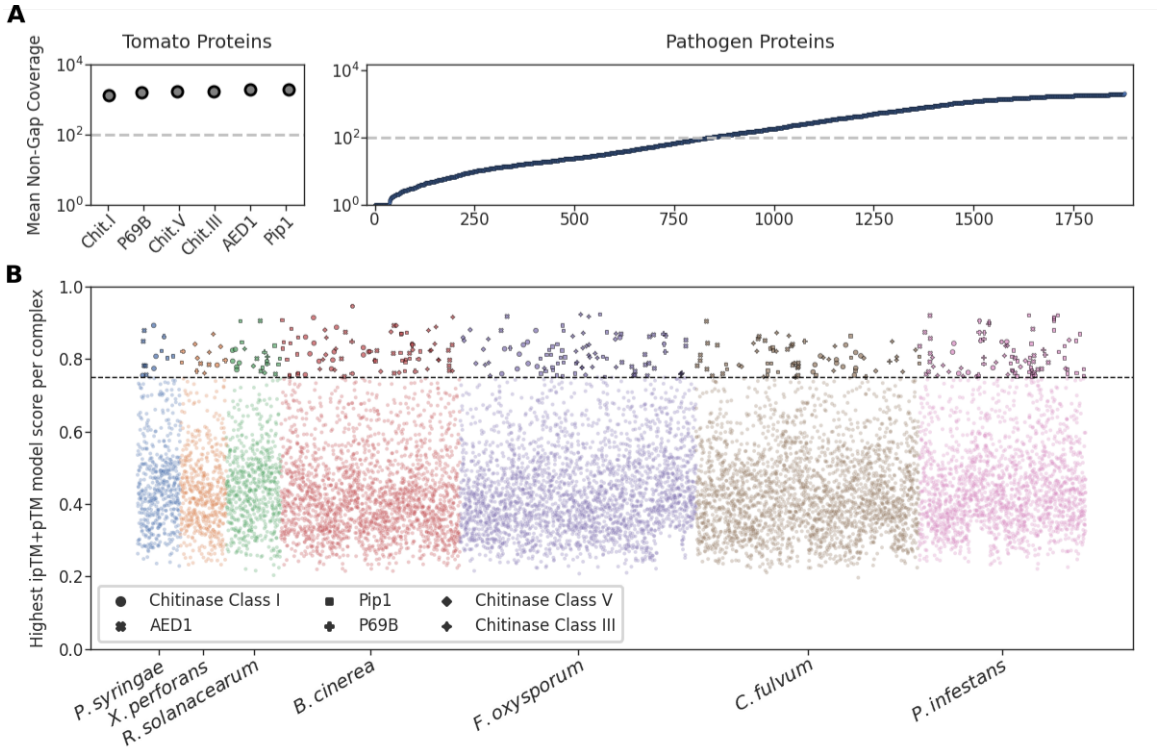


Figure 53: Modelling 11,274 inhibitor-enzyme complexes identifies high-scoring protein complexes. (**A**) Mean non-gap amino acid coverage in protein multi sequence alignments (MSAs) used to predict protein complexes between six tomato proteins and 1,879 pathogen-derived proteins. Grey dashed horizontal lines indicate a mean non-gap MSA depth of 100. (**B**) Manhattan plot showing the model with the highest ipTM+pTM score for each of the 11,274 modelled complexes. The black horizontal dashed line indicates the cut off ipTM+pTM value of 0.75.

all host proteins was sufficient. For pathogen-derived proteins, 54.98% of the 1,879 pro-

teins showed a mean non-gap MSA depth above the MSA saturation depth of 100 (**Figure 53A**). Since the highest-scoring protein complex of the known inhibitor-enzyme pair Avr2 and Rcr3 (highest ipTM+pTM score of 0.73) scored close to the ipTM+pTM threshold of 0.75 despite very low mean non-gap MSA depth of Avr2 (5.83), we decided to retain all 1,879 effector proteins. From the *Sl*C14-*Pto*Cip1 positive control we learned that not all five predicted models for each protein complex score similarly high. We thus decided to only consider the highest-scoring model for each of the 11,274 protein pairs. We identified 376 (3.3% of screened) protein pairs containing at least one model with an ipTM+pTM $\geq$ 0.75 (**Figure 53B**). We identified high-scoring protein pairs for all six tomato proteins: class I chitinase (51), class III chitinase (69), class V chitinase (27), P69B (45), Pip1 (106) and AED1 (78). Of these 376 protein pairs, 140 of pathogen-derived proteins (37.2%) had a mean non-gap MSA depth below information saturation of 100, indicating that many high-scoring complexes were predicted despite low sequence information of these 140 effectors.

5.3.3 Manual scoring of 376 protein complexes identifies putative novel apoplastic hydrolase inhibitors

Next, we manually evaluated 376 predicted protein complexes to identify putative novel enzyme-inhibitor complexes. We automated the visualisation of protein complexes with a custom-made python script to display the host target as surface representation in grey with highlighted active site to quickly evaluate the impact of effector binding. Models of effector-host target complexes were evaluated and classified into four categories: Category 0 contained misfolded, nonsense protein complexes; category 1 contained protein complexes with protein-protein interactions not obstructing the active site; category 2 contained complexes with protein-protein interactions obstructing the active site, but without secondary structural support for the putative inhibitory protein segment; and category 3 contained complexes with protein-protein interactions blocking the active site and with structural support for the putative inhibitory segment such as beta sheets and/or disulfide bridges.

First, we analysed the distribution of screened protein complexes in the four different scoring categories. Only 19 of 376 protein complexes (5.1%) still fell into the misfolded (0) category, ranging from 2.9% for class III chitinase- to 7.4% for class V chitinase complexes (**Figure 54A**). These protein complexes showed intertwined amino acid chains or prox-

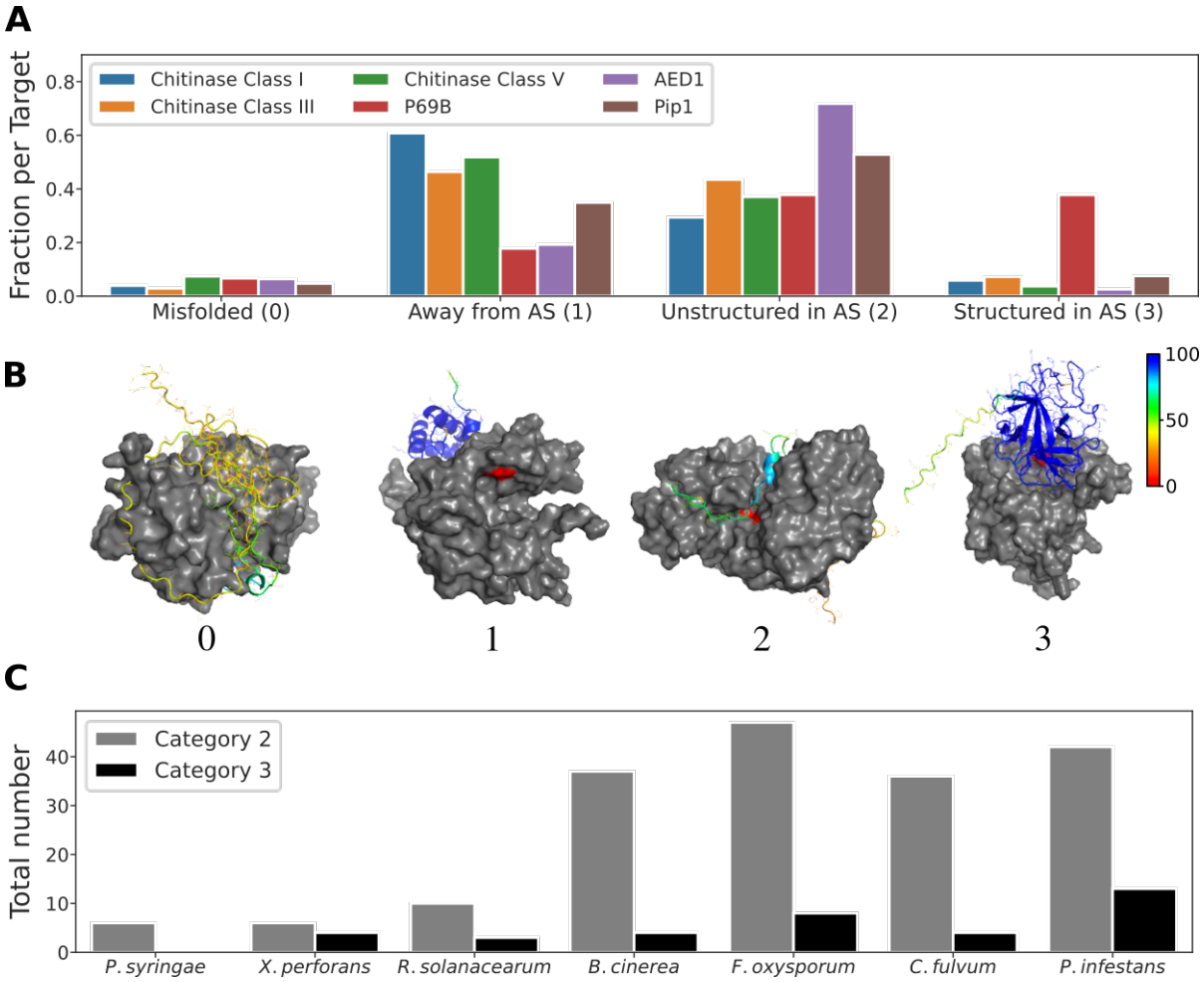


Figure 54: Manual classification of protein complexes. (**A**) Manual screening of 376 protein complexes and grouping of each into misfolded, nonsense complexes (category 0), interactions between putative inhibitor and plant tomato PR protein away from the target's active site (AS) (category 1), interactions between the putative inhibitor and tomato PR protein near the active site, but no secondary structural support for the putative inhibitory protein segment (category 2) and interactions near the tomato protein's active site and the putative inhibitor with secondary structural support for the potential inhibitory protein segment (category 3). (**B**) Protein complexes marked as 0, 1, 2 and 3 each represent examples of their respective scoring categories. (**C**) Number of category 2 and 3 protein complexes including proteins from each of the seven plant pathogens.

imity of residues that do not appear biologically possible (**e.g. Figure 54B 0**). These examples highlight room for improvement of protein complex prediction by AlphaFold2

multimer. Category 1 contains putative protein-protein interactions where the active site is not blocked by the effector. 137 of 376 protein complexes (36.4%) fell in this category, ranging from 17.8% for P69B to 60.8% for class I chitinase complexes. These predicted protein complexes might include allosteric hydrolase inhibitors (e.g. **Figure 54B 1**). The concept of allosteric inhibition at apoplast plant-pathogen interfaces is intriguing, but was not pursued further. Category 2 contains proteins with putative protein-protein interactions blocking the active site, but without further structural support for the interacting segment that would stabilise the putative protein-protein interaction. This category contained 184 of the 376 protein complexes (48.9%) and ranged from 29.4% for class I chitinases- to 71.8% for AED1 complexes. These putative protein complexes may include possible substrates for proteases (e.g. **Figure 54B 2**). The question whether AlphaFold2 multimer can accurately predict enzyme-substrate interactions is intriguing, but was not further pursued in this work. Category 3 contains putative protein-protein interactions blocking the active site with additional structural support for the interacting protein segment. This category contained 36 of the 376 protein complexes (9.6%) and ranged from 2.6% for AED1- to 37.8% for P69B complexes (e.g. **Figure 54B 3**). 13 of these predicted protein complexes involve cystatin-like and kazal-like protein interactions with Pip1 and P69B, respectively. These protein pairs were removed from downstream analyses to focus on novel putative inhibitors at the plant pathogen interface. The remaining 23 protein pairs included all six tomato proteins: class I chitinase (3), class III chitinase (5), class V chitinase (1), P69B (9), Pip1 (3) and AED1 (2). We identified most novel putative inhibitors for P69B which is the most abundant hydrolase in AF from Bion-induced tomato (**chapter 3 Figure 34**). We identified more category 2 than 3 protein complexes for all plant pathogens and most category 3 proteins pairs (13) include proteins from *P.infestans* (**Figure 54C**). These putative hydrolase inhibitors were solely identified based on their high predicted ipTM+pTM score in complex with apoplastic tomato hydrolases their manually assigned category 3 score of the predicted complex.

5.3.4 Expression of putative hydrolase inhibitors

To substantiate our hypothesis of the 23 putative novel hydrolase inhibitors acting at the plant-pathogen interface, we analysed *in planta* gene expression data of the seven plant pathogens. Therefore, we downloaded *in planta* RNAseq data for five of the seven plant pathogens (*P.syringae* pv. *tomato, R.solanacearum, B.cinerea, F.oxysporum* f. sp. *lycopersici* and *C.fulvum*). Since expression data was collected from host tissue, we took a metagenomics approach and classified sequencing reads to discard all reads marked as host-derived. Then, we identified expressed genes by aligning reads to all nucleotide sequences derived from structural annotations of the relevant pathogen. This way, we identified pathogen-derived genes that are expressed during infection. Since the RNAseq data sets were heterogeneous for host species, infected tissue, sampling time points and number of replicates, we classified a gene as expressed during infection if transcripts were detected at ≥ 2 transcripts per million (tpm). To put our threshold of 2 tpm into context, the European Bioinformatics Institute's Gene Expression Atlas defines its lower cut off at 0.5 tpm. For a more detailed description, see Materials and Methods 5.2.6.

Of our 23 novel putative hydrolase inhibitors, we detected *in planta* gene expression for 11 genes, ranging from 2.44 to 717.5 tpm (**Figure 55**). Interestingly, APP91304.1 is encoded on chromosome 14 of *F.oxysporum* f. sp. *lycopersici* 4287 and had been added manually to coding sequences as it was not included in the reference structural annotation, showed the highest level of *in planta* gene expression of all putative novel hydrolase inhibitors. Four putative hydrolase inhibitors are from *X.perforans* for which we could not find *in planta* RNAseq data. Eight putative inhibitors did not show robust *in planta* gene expression. Even though this could be due to sampling time points of expression data, we decided to exclude all eight putative hydrolase inhibitors for which we could not detect gene expression during infection despite *in planta* RNAseq data.

We continued with putative hydrolase inhibitors expressed *in planta* and included the four proteins without available expression data. This data set combined 15 putative hydrolase inhibitors.

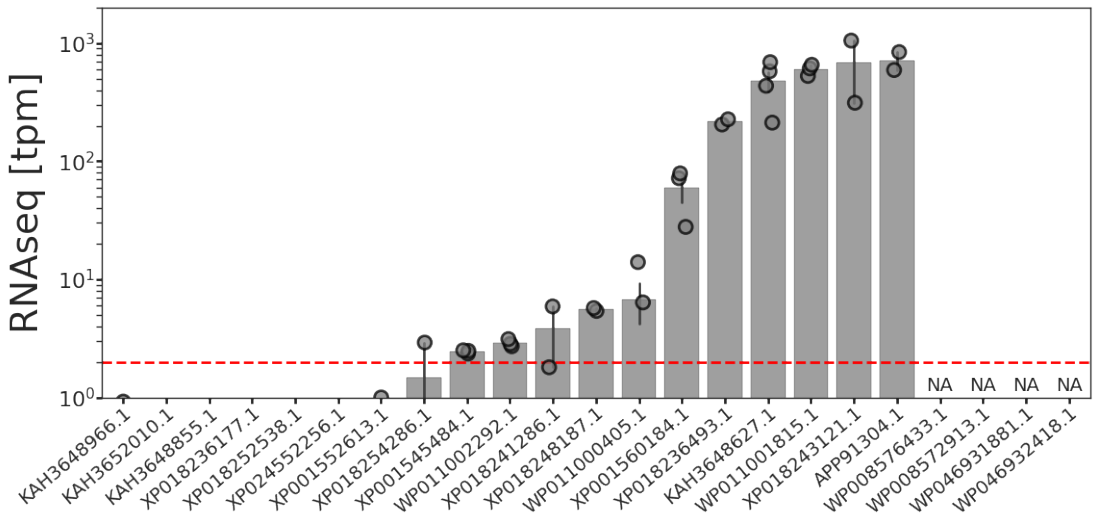


Figure 55: *In planta* expression evidence of novel hydrolase inhibitors. *In planta* expression was detected using published RNAseq data, mapped onto structural annotations used in this chapter. An overview of RNAseq data sets can be found in **Table 13**. APP91304.1 was added manually to the structural annotation of *F.oxysporum* f. sp. *lycopersici*. Each data point represents a replicate and bars indicate the mean and standard error of means. The horizontal, red dotted line indicates the threshold of 2 transcripts per million [tpm] for a gene to be considered expressed. The last four proteins marked with 'NA' are encoded by *X.perforans* for which no *in planta* expression data was found.

5.3.5 Conservation of putative hydrolase inhibitors

Next, to compare the 15 putative hydrolase inhibitors to homologs in other plant pathogens, we clustered proteomes of 37 prominent plant pathogens into orthogroups using Orthofinder2.0 (Emms and Kelly 2019). The plant pathogens were selected based on scientific and agricultural importance and included bacteria, oomycetes and fungi (Dean et al. 2012; Mansfield et al. 2012; Kamoun et al. 2015). The phylogenetic tree generated by orthofinder supported the division into a bacterial, oomycete and fungal clade (**Figure 56**). Furthermore, we detected a clear distinction between *Ralstonia, Xanthomonas* and *Pseudomonas* bacterial subclades as well as a *Phytophthora* oomycete subclade (**Figure 56**). Some putative hydrolase inhibitors are species-specific such as *Ralstonia solanacearum*'s WP011000405 and WP011002292.1, *Cladosporium fulvum*'s KAH3648627.1 and *Fusarium oxysporum*'s XP018243121.1 and APP91304.1. Other putative hydrolase inhibitors are subclade-specific such as WP046932418.1, WP046931881.1 and WP008572913.1 in

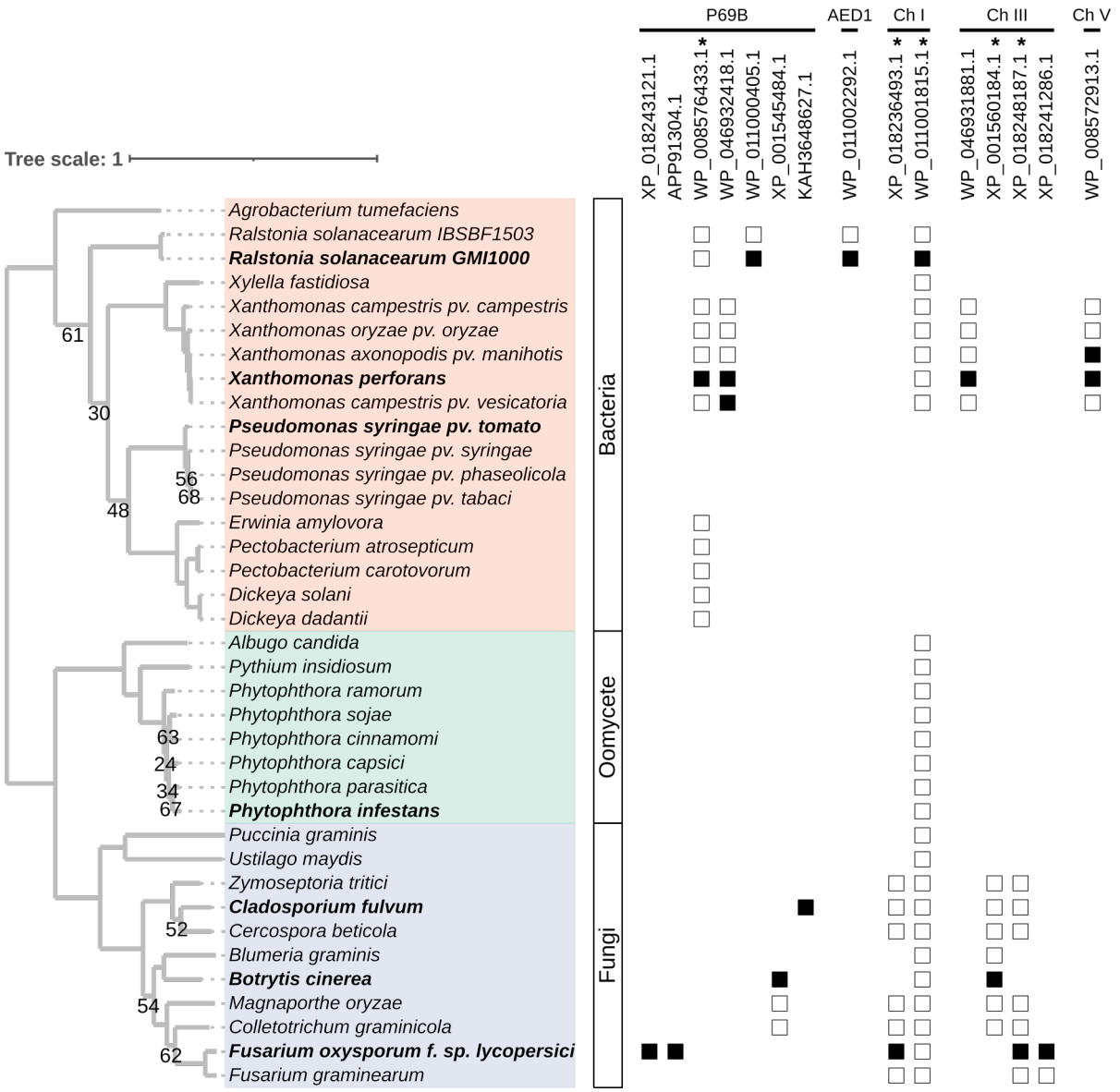


Figure 56: Conservation of putative hydrolase inhibitors. A species phylogenetic tree derived from 37 plant-pathogen proteomes, inferred and rooted from whole-proteome data by Orthofinder2.0 (Emms and Kelly 2019). Plant pathogens involved in the screen to identify novel hydrolase inhibitors are highlighted in bold. Percentage support values below 70 are shown. Filled squares adjacent to leaves represent putative hydrolase inhibitors and empty squares indicate homologs to the putative inhibitor in respective proteomes. Asterisks indicate orthogroup with potentially conflicting annotations of homologs as hydrolase inhibitors. Ch I, Ch III and Ch V stand for class I, III and V chitinase respectively.

X.perforans. Interestingly, WP046932418.1 from *X.perforans* is also found in *X.campestris* pv. *vesicatoria*'s and WP008572913.1 from *X.perforans* is also found in *X.axonopodis pv. manihotis*' proteome.

The remaining five putative hydrolase inhibitors (WP008576433.1, XP018236493.1, WP 011001815.1, XP001560184.1 and XP018248187.1) have homologs in more distant plant

pathogens. Some of these pathogens had been included in the AlphaFold2 multimer screen, but these homologs had not been identified as putative hydrolase inhibitors. This raises two central questions: First, why did AlphaFold2 multimer not predict good complexes for these homologs? Second, should we deselect any of the five proteins if their homologs had not been selected? Therefore, analysing homologs could highlight inconsistencies in our approach. These five different inhibitor candidates are discussed in the next section to evaluate the prediction of complexes by AlphaFold2 multimer.

5.3.6 Novel hydrolase inhibitors predicted by AlphaFold2 multimer

5.3.6.1 WP008576433.1

Based on AlphaFold2 multimer's ipTM+pTM score, WP008576433.1 is a putative P69B inhibitor from *X.perforans* and it is annotated as an entericidin A/B family lipoprotein and has one apoplastic homolog in *R.solanacearum* (WP011004496.1). While both proteins in complex with P69B have low mean scores, WP008576433.1 showed the higher best ipTM+pTM score, just exceeding the threshold of 0.75 while WP011004496.1 scored below the threshold (**Table 16**). *R.solanacearum*'s WP011004496.1 was thus not evaluated as a putative hydrolase inhibitor due to its low ipTM+pTM scores. WP008576433.1 has the higher mean non-gap MSA depth at 96.22, compared to 25.18 for WP011004496.1, which could explain why predicted ipTM+pTM scores were low for this homolog (Jumper et al. 2021). Unfortunately, no RNAseq data was available for *X.perforans*, but *R.solanacearum*'s WP011004496.1 was detected during infection at 149.2 tpm.

Table 16: WP_008576433.1 secreted homologs screened

		ipTM+pTM			
Species	Protein	All Models	Best Model	MSA depth[1]	Expression [tpm]
X.perforans	WP008576433.1	0.65 ± 0.06	**0.77**	96.22	-
R.solanacearum	WP011004496.1	0.51 ± 0.02	0.54	25.18	$\mathbf{149.2 \pm 19.6}$

According to NCBI's functional annotation, WP008576433.1 is described as an entericidin A/B family lipoprotein. This could indicates that WP008576433.1 might not be a novel hydrolase inhibitor. Entericidin A/B genes encode two small, paralogous lipoproteins implicated in bacterial cell lysis (Bishop et al. 1998). Entericidin B has a bacteriolytic

function while entericidin A is the antidote, suppressing entericidin B-mediated cell lysis (Bishop et al. 1998). Despite WP008576433.1 scoring just above the ipTM+pTM threshold of 0.75 and showing a good putative enzyme-inhibitor complex, it was not further considered based on its functional annotation as an entericidin A/B family protein.

5.3.6.2 XP018236493.1

Based on AlphaFold2 multimer's ipTM+pTM score, XP018236493.1 is a putative chitinase class I inhibitor from *F.oxysporum* f. sp. *lycopersici* (*Fol*). The relevant orthogroup contains two additional apoplastic proteins of *Fol* and three apoplastic proteins of *C.fulvum*. *C.fulvum* KAH3643982.1 has a low MSA depth of 19 which is why it will

Table 17: XP018236493.1 secreted homologs screened

		ipTM+pTM			
Species	Protein ID	All Models	Best Model	MSA depth[1]	Expression [tpm]
F.oxysporum	XP018236493.1	0.85 ± 0.03	**0.88**	**355.35**	$\mathbf{216.06 \pm 11.03}$
F.oxysporum	XP018249861.1	0.27 ± 0.02	0.29	**351.91**	$\mathbf{358.87 \pm 19.18}$
F.oxysporum	XP018258017.1	0.27 ± 0.04	0.33	**321.89**	0.99 ± 0.99
C.fulvum	KAH3648558.1	0.33 ± 0.12	0.57	**276.90**	$\mathbf{3881.44 \pm 1832.44}$
C.fulvum	KAH3656077.1	0.28 ± 0.03	0.31	**124.79**	$\mathbf{88.94 \pm 57.64}$
C.fulvum	KAH3643982.1	0.35 ± 0.04	0.39	19.16	$\mathbf{12.27 \pm 8.57}$

not be considered for the rest of the discussion (**Table 17**). Overall, the high mean non-gap MSA depths of the remaining *C.fulvum* and all Fol proteins instil confidence in the quality of AlphaFold2 multimer's predicted protein complexes. Turning to our *in planta* expression data, all three *C.fulvum* proteins were expressed during infection and also two *Fol* proteins (XP018236493.1 and XP018249861.1) were expressed during infection. The third *Fusarium* protein, XP018258017.1, fell below the 2 tpm minimum expression threshold. Since mean non-gap MSA depth, ipTM+pTM scores and expression data could not explain why XP018236493.1 is the only protein in the orthogroup identified as a novel putative hydrolase inhibitor, we analysed the interactive regions of the six proteins further. KAH3643982.1 is again noticeably different to the other five analysed homologs in the orthogroup and does not have the two inhibitory loops present in the other proteins (**Figure 57B**). The putative inhibitory domains of XP018236493.1 mostly consist of two loops in the active site (**Figure 57A**). The first loop is not fully conserved

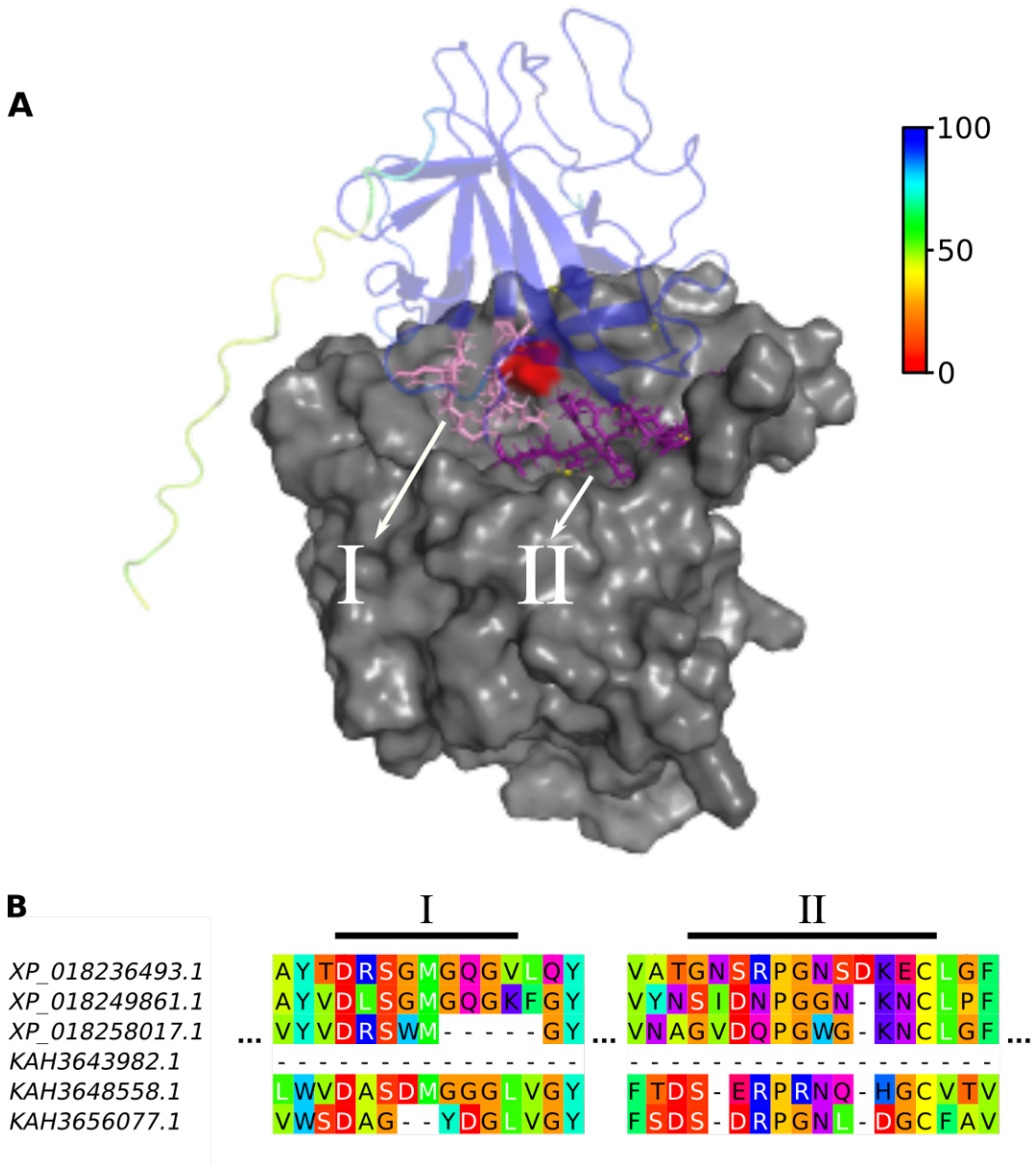


Figure 57: Comparison of putative inhibitory loops between homologs of the Chitinase class I inhibitor XP018236493.1. (**A**) Chitinase class I in complex with XP018236493.1. The inhibitor is predominantly coloured based on plDDT values of each amino acid which is indicated by the color bar. The two inhibitory loops I and II, interacting with the SBG are coloured in pink and purple respectively. The chitinase is depicted in grey with the catalytically active glutamate coloured red. (**B**) The MSA of the six secreted homologs comprising the two inhibitory loops marked above the sequences.

in the homologs, but XP018249861.1 shows the highest sequence similarity with only two amino acid substitutions compared to XP018236493.1 (**Figure 57B**). The second putative inhibitory loop is not conserved between the homologs and XP018249861.1 has eight amino acid substitution in the second loop. The sequences could explain why homologs of

XP018236493.1 could not display specific interactions with the class I chitinase because the two loops at the interaction interface are poorly conserved.

In conclusion, this orthogroup contains six apoplastic proteins of *C.fulvum* and *Fol* five of which show high mean non-gap MSA depth to predict high confidence scores with the class I chitinase. Likely due to severely altered putative inhibitory loops, only XP018236493.1 was predicted to interact with the class I chitinase. Taken together, ipTM+pTM scores, mean non-gap MSA depth, RNAseq expression and sequence comparisons with homologs only strengthen the selection of XP018236493.1 as a putative class I chitinase inhibitor.

5.3.6.3 WP011001815.1

Based on AlphaFold2 multimer's ipTM+pTM score, WP011001815.1 is a putative class I chitinase inhibitor from *R.solanacearum*. Its orthogroup spans bacteria, oomycetes and fungi and it is functionally annotated as a YSC84-related protein. In its orthogroup, WP011001815.1 is the only small protein from the seven tested plant pathogens predicted to be apoplastic. While one homolog in each *X.perforans* (WP008577761.1) and *P.infestans* (XP002894766.1) were predicted to be secreted, neither was predicted by ApoplastP to remain apoplastic. Homologs could thus not be identified as putative hydrolase inhibitors since they had not been included in AlphaFold2 multimer modelling. WP011001815.1 is highly expressed during infection (599.33 tpm) and its mean non-gap

Table 18: WP011001815.1 homologs screened

		ipTM+pTM			
Species	Protein ID	All Models	Best Model	MSA depth[1]	Expression [tpm]
R.solanacearum	WP011001815.1	0.50 ± 0.19	**0.83**	**1798.12**	$\mathbf{599.33 \pm 53.24}$

MSA depth is high (1798.12). The low mean scores of the complexes with WP011001815.1 stand out however. Its functional annotation in NCBI as a YSC84-related protein does not implicate it in apoplastic plant-pathogen interactions. YSC84 proteins are actin-binding proteins involved in organising the cytoskeleton (see definition of IPR007461).

In conclusion, homologs were not tested as putative hydrolase inhibitors since none of them had been predicted to be small apoplastic proteins. Further functional annotation

makes it unlikely that WP011001815.1 is a novel class I chitinase inhibitor, but rather a false positive in our AlphaFold2 multimer screen. WP011001815.1 was not further considered as a novel putative hydrolase inhibitor.

5.3.6.4 XP001560184.1

Based on AlphaFold2 multimer's ipTM+pTM score, XP001560184.1 is a putative class III chitinase inhibitor from *B.cinerea* and only *C.fulvum* contains one apoplastic homolog, KAH3648475.1. KAH3648475.1 has a low mean non-gap MSA depth of 9.48 and consistently low predicted ipTM+pTM scores. XP001560184.1 on the other hand has high mean non-gap MSA depth of 296, a moderate average ipTM+pTM score of 0.63 with high variability and a high-scoring 0.88 best model. XP001560184.1 is expressed in tomato during infection, KAH3648475.1 is not.

Table 19: WP014616014.1 secreted homologs screened

		ipTM+pTM			
Species	Protein ID	All Models	Best Model	MSA depth[1]	Expression [tpm]
B.cinerea	XP001560184.1	0.63 ± 0.24	**0.88**	**296.24**	$\mathbf{59.56 \pm 22.65}$
C.fulvum	KAH3648475.1	0.29 ± 0.01	0.30	9.48	0

In conclusion, the homolog KAH3648475.1 was likely not identified as a novel putative hydrolase inhibitor due to its low mean non-gap MSA depth to accurately predict ipTM+pTM scores of resulting complexes. Furthermore, KAH3648475.1 is not expressed during infection of tomato leaves. XP001560184.1 remains a novel putative hydrolase inhibitor since no contradicting evidence was uncovered.

5.3.6.5 XP018248187.1

Based on AlphaFold2 multimer's ipTM+pTM score, XP018248187.1 is a putative class III chitinase inhibitor from Fol that only has one homolog in *C.fulvum* which was not predicted to be secreted and was therefore not tested by AlphaFold2 multimer. In conclusion, lack of secretion explains why no homolog had been included in our modelling approach to identify putative hydrolase inhibitors. Analysing homologs did not reveal evidence against the identification of XP018248187.1 as a putative hydrolase inhibitor.

Table 20: XP018248187.1 secreted homologs screened

		ipTM+pTM			
Species	Protein ID	All Models	Best Model	MSA depth[1]	Expression [tpm]
F.oxysporum	XP018248187.1	0.547 ± 0.277	**0.916**	**171.591**	$\mathbf{5.56 \pm 0.16}$

Upon combining analyses of putative novel inhibitor proteins and their homologs, we confidently resolved potential conflicting evidence for three putative inhibitors XP018236493.1, XP001560184.1 and XP018248187.1. WP008576433.1 was discarded based on its functional annotation as part of the suicidal, toxin-antidote pair Entericind A/B and WP01 1001815.1 was discarded based on conflicting functional annotation as a YSC84-related protein.

The ten remaining pathogen-derived proteins were identified as putative hydrolase inhibitors based on high-scoring predicted protein complexes with cognate tomato hydrolases and detected *in planta* expression. These ten putative hydrolase inhibitors encoded two chitinase class III-, one chitinase class V-, six P69B- and one AED1 inhibitor. The two putative chitinase class III inhibitors WP046931881.1 and XP018241286.1 were produced by *X.perforans* and *Fol* respectively. WP046931881.1 is functionally annotated as a membrane-bound lysozyme inhibitor of c-type lysozymes (MliC) and XP018241286.1 is a hypothetical protein only. MSA depth of XP018241286.1 was lower at 62.22 compared to 141.47 for WP046931881.1 (MliC) and the highest-scoring model in complex with the class III chitinases of both had an ipTM+pTM score of 0.87. Expression during infection of *Fol* XP018241286.1 was low (3.85 tpm) and no expression data was available for *X.perforans*. The putative chitinase class V inhibitor WP008572913.1 was produced by *X.perforans* and encoded a hypothetical protein. Its MSA depth was above information of saturation at 107.08, the best model in complex with the class V chitinase had an ipTM+pTM score of 0.83 and no expression data was available. The six putative P69B inhibitors were detected in *X.perforans* (WP046932418.1), *R.solanacearum* (WP011000405.1), *B.cinerea* (XP001545484.1), *C.fulvum* (KAH3648627.1) and *Fol* (XP018243121.1 and APP91304.1). The two *Fol*-derived putative inhibitors encoded a trypsin-inhibitor like protein (XP01 8243121.1) and secreted in xylem 15 (Six15, APP91304.1) (Simbaqueba et al. 2021).

Both proteins are highly-expressed during infection (>600 tpm) and trypsin inhibitor-like protein had the deepest MSA coverage of the 13 candidates (815.69) and Six15 the lowest (3.88). Best models of both had high ipTM+pTM scores at 0.83 (Six15) and 0.87 (trypsin inhibitor-like) in complexes with P69B . The fungal-derived P69B inhibitors XP001545484.1 (*B.cinerea*) and KAH3648627.1 (*C.fulvum*) are annotated as phosphatidyl ethanolamine-binding protein in eukaryotes (PEBP Euk.) and extracellular protein 36-1 (Ecp36-1) (Mesarich et al. 2018). Expression during infection was high for Ecp36-1 (479.16 tpm) and low for PEBP Euk. (2.44 tpm) while PEBP Euk. had high MSA coverage (238.30) and Ecp36-1 had the second lowest MSA coverage of all 13 candidate inhibitors (13.43). The best model in complex with P69B of Ecp36-1 only had an ipTM+pTM score of 0.75 and the best model of PEBP Euk. received an ipTM+pTM score of 0.82. The putative bacterial P69B inhibitors WP011000405.1 (*R.solanacearum*) and WP046932418.1 (*X.perforans*) are both annotated as hypothetical proteins. WP011000405.1 expression during infection is low (6.78 tpm) and no expression data for *X.perforans* was available and the MSA depths of both proteins exceed saturation of information. *R.solanacearum*'s WP011000405.1 highest scoring model in complex with P69B had an ipTM+pTM score of 0.81 and the highest-scoring model for WP046932418.1 from *X.perforans* was 0.82. For AED1 we detected a single putative inhibitor WP011002292.1 from *R.solanacearum*, which was a lowly-expressed (2.89 tpm) hypothetical proteins with deep MSA coverage (576.77) and its highest scoring model in complex with AED1 had an ipTM+pTM score of 0.76. Overall, we noticed deep MSA coverage was above saturation of information for 10 of the remaining 13 protein candidates (76.92%). In addition to our identification of these 13 proteins as putative hydrolase inhibitors based on their ipTM+pTM scores, manually assigned complex scores and *in planta* expression, two pathogen-derived proteins were already predicted to be hydrolase inhibitors based on their functional annotations. These included Membrane-bound lysozyme-inhibitor of c-type (WP046931881.1) and a putative trypsin inhibitor (APP91304.1). All 13 putative hydrolase inhibitors are summarised in **Table 21**. Finally, we analysed putative target specificity of each of the 13 final putative hydrolase inhibitors. To show the predicted selectivity of the putative inhibitors,

Table 21: Summary of 13 putative hydrolase inhibitors

Proteins	Species	Protein ID	Funct. annot.	ipTM+pTM		MSA depth	Expression [tpm]
				Mean	Best		
Chit. I	*F.oxysporum*	XP018236493.1	CW PhiA[0*]	0.85 ± 0.03	0.88	355.35	216.06 ± 11.03
Chit. III	*X.perforans*	WP046931881.1	MliC[1] Fam.	0.76 ± 0.20	0.87	141.47	-
Chit. III	*B.cinerea*	XP001560184.1	Hypoth. Prot.[2]	0.63 ± 0.24	0.88	296.23	59.56 ± 22.65
Chit. III	*F.oxysporum*	XP018248187.1	FolNIS1[3*]	0.55 ± 0.28	0.92	171.59	5.56 ± 0.16
Chit. III	*F.oxysporum*	XP018241286.1	Hypoth. Prot.	0.43 ± 0.23	0.87	62.22	3.85 ± 2.05
Chit. V	*X.perforans*	WP008572913.1	Hypoth. Prot.	0.49 ± 0.18	0.83	107.08	-
P69B	*X.perforans*	WP046932418.1	Hypoth. Prot.	0.64 ± 0.16	0.82	213.55	-
P69B	*R.solanacearum*	WP011000405.1	Hypoth. Prot.	0.59 ± 0.12	0.81	614.55	6.78 ± 5.71
P69B	*B.cinerea*	XP001545484.1	PEBP Euk.[4]	0.57 ± 0.16	0.82	238.30	2.44 ± 0.07
P69B	*C.fulvum*	KAH3648627.1	Ecp36-1[5]	0.55 ± 0.11	0.75	13.43	479.16 ± 178.70
P69B	*F.oxysporum*	XP018243121.1	Trypsin inh.*	0.79 ± 0.05	0.87	815.69	682.91 ± 369.84
P69B	*F.oxysporum*	APP91304.1	Six15[6]	0.69 ± 0.12	0.83	3.88	717.49 ± 124.45
AED1	*R.solanacearum*	WP011002292.1	Hypoth. Prot.	0.67 ± 0.10	0.76	576.77	2.89 ± 0.17

[0] Cell wall protein PhiA. [1] Membrane-bound lysozyme-inhibitor of c-type. [2] Hypothetical protein. [3] *F.oxysporum* f. sp. *lycopersici* Necrose-inducing secreted protein 1. [4] Phosphatidyl Ethanolamine-Binding Protein in eukaryotes. [5] Extracellular Protein 36-1. [6] Secreted in Xylem 15. * Functional annotation based on Blastp searches.

we extracted the ipTM+pTM value of the highest scoring model between the final 13 putative inhibitors and the six tomato proteins. Only a single putative enzyme-inhibitor interaction for each complex had manually scored as category 3 (**Figure 58**). Putative

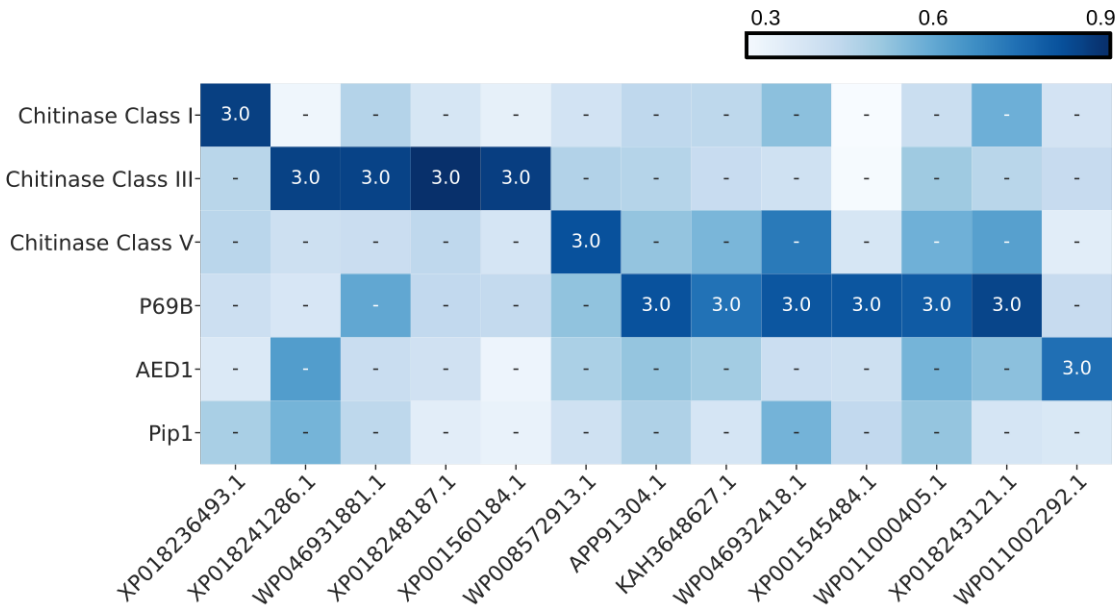


Figure 58: Putative target specificity of 13 putative hydrolase inhibitors. Colours depict the highest ipTM+pTM score of five predicted models of protein complexes. The values display the manually scored category of each complex.

inhibitor-enzyme models of these 13 putative hydrolase inhibitors can be found in **Figures 59 and 60**.

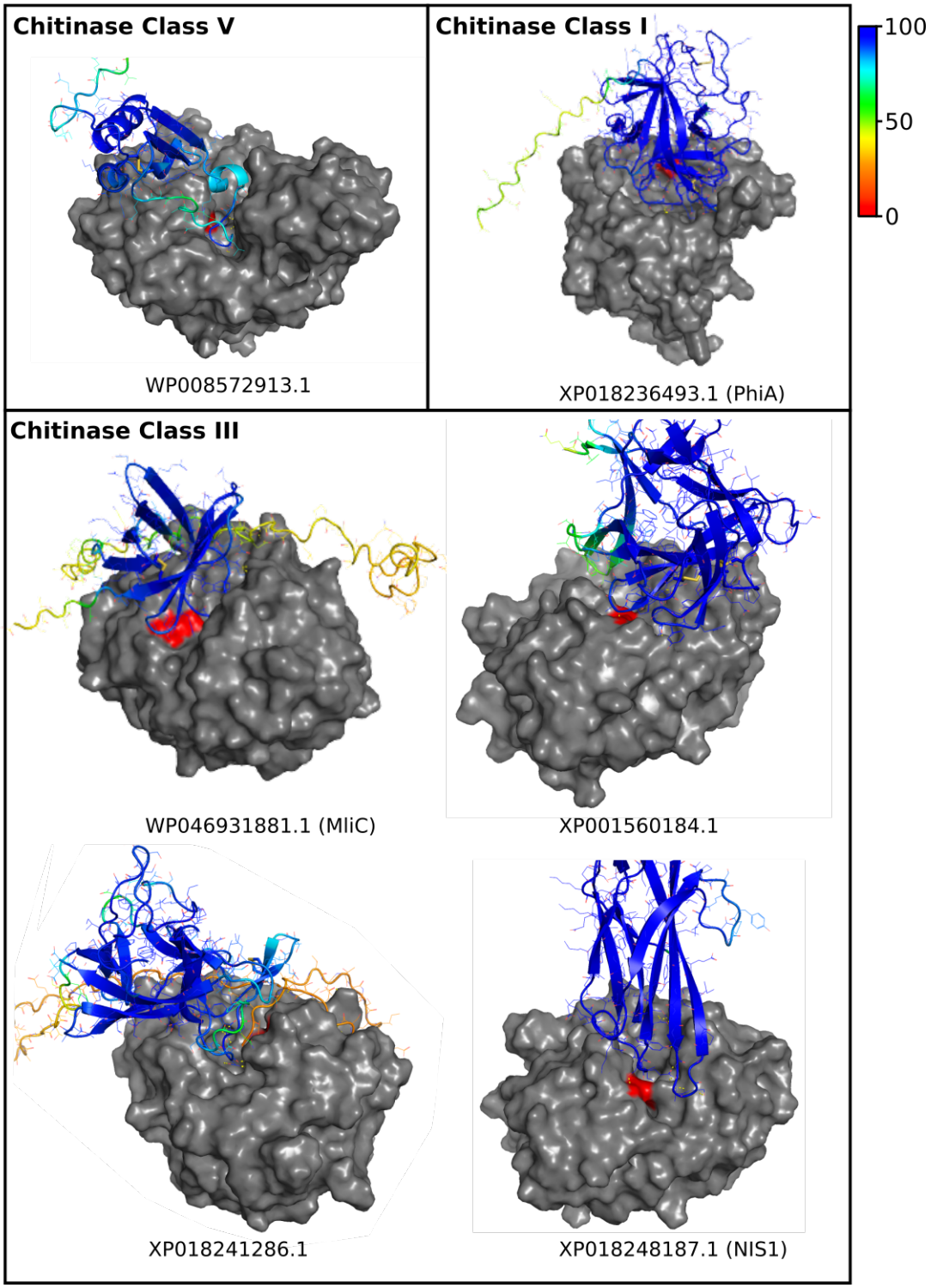


Figure 59: Overview of putative inhibitor-chitinase complexes. Target proteins are coloured grey, their catalytically active residue is coloured red. Putative pathogen-derived hydrolase inhibitors are coloured based on their plDDT scores.

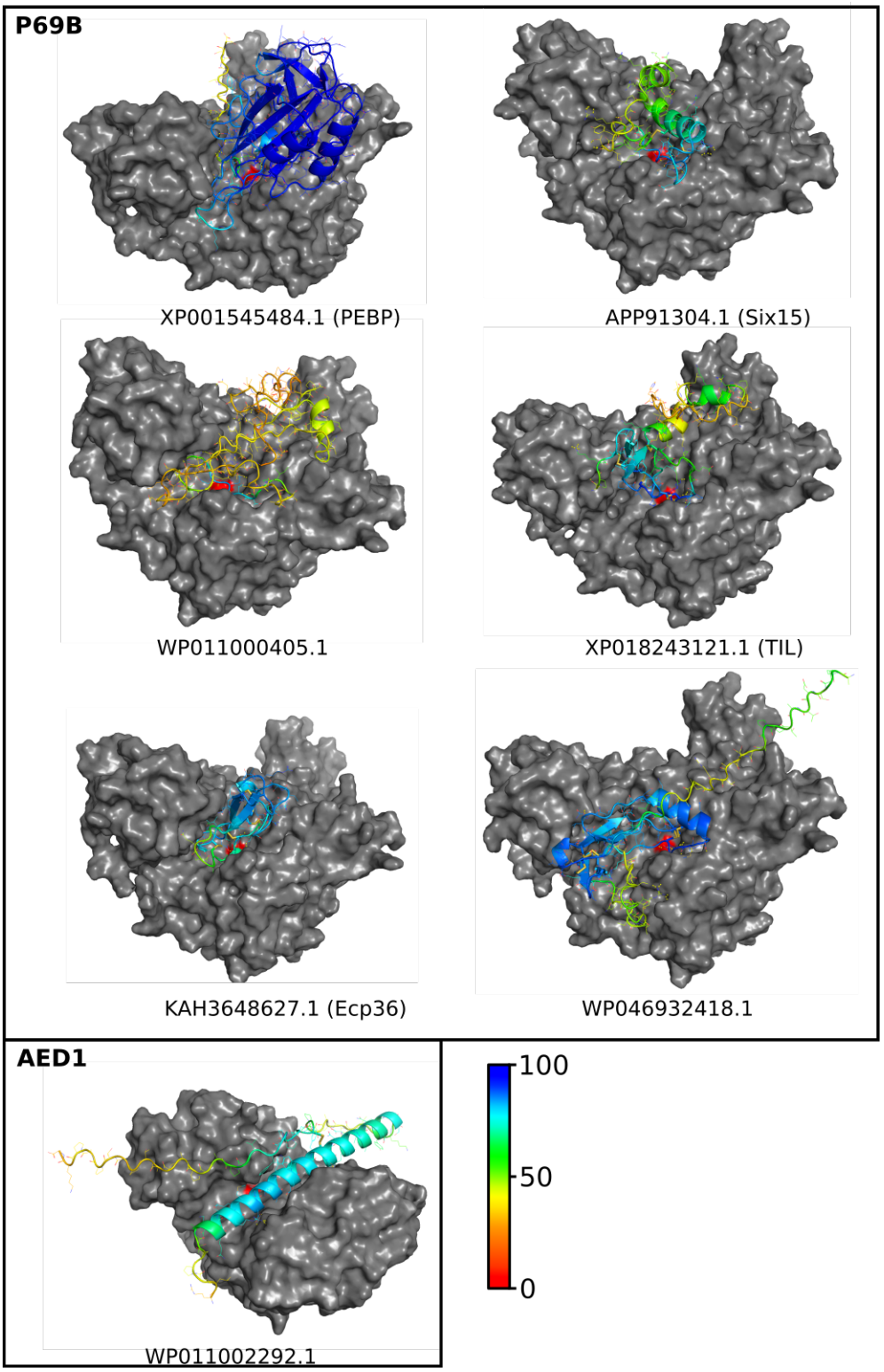


Figure 60: Overview of putative inhibitor-protease complexes. Target proteins are coloured grey, their catalytically active residue is coloured red. Putative pathogen-derived hydrolase inhibitors are coloured based on their plDDT scores.

5.4 Discussion

In this chapter, we established AlphaFold2 multimer as a high-throughput tool to predict and identify interspecies inhibitor-enzyme complexes from the host-pathogen interface. First, we provided evidence that AlphaFold2 multimer-derived scores can identify known host-pathogen protein complexes from the apoplastic plant-pathogen interface. Second, we computed 11,274 protein complexes between tomato proteins and predicted apoplastic proteins of seven plant pathogens and identified 376 high-scoring complexes. Third, upon manual screening of the 376 high-scoring protein complexes, we identified a subset of 23 putative novel inhibitor-enzyme complexes in addition to several complexes with cystatin- and kazal-like proteins. Finally, we embedded the 23 putative enzyme inhibitors in a broader biological context by analysing *in planta* expression data and comparing the distribution of these putative inhibitors in other plant pathogens. This last step provided a reduced list of 13 novel putative hydrolase inhibitors, likely acting at the plant-pathogen interface in the tomato apoplast.

5.4.1 AlphaFold2 models experimentally validated protein complexes at the apoplastic plant-pathogen interface with high predicted quality

We showed that AlphaFold2 multimer can predict known inhibitor-enzyme complexes between plant- and pathogen-derived proteins. We selected 28 experimentally validated protein complexes from the apoplastic plant-pathogen interface and evaluated their predicted 3D models based on their predicted superposition ipTM+pTM scores. No comparisons between our predicted complexes with experimentally resolved complexes of the 28 inhibitor enzyme pairs were possible, since no complexes were publicly available.

While all tested host-derived apoplastic hydrolases had mean non-gap MSA depths above 1,000, ten of the 28 pathogen-derived proteins had a mean non-gap MSA depth below 15. None of these ten resulted in predicted ipTM+pTM scores above 0.75 (**Figure 49B**). In our small data set, we found that for sufficiently high mean non-gap MSA depths (>100), recall of positive controls was high for known inhibitor-enzyme complexes (77.78%) and even higher for inhibitor-protease complexes (92.86%) (**Figure 49B**). Furthermore, speci-

ficity at high mean non-gap MSA depths was high (**Figure 51**).

Applying AlphaFold2 multimer to predict inhibitor-enzyme complexes at host-pathogen interfaces was entirely novel and two major conclusions can be drawn from this section. First, while all protein complexes used as positive controls had been extensively characterised experimentally, none of the protein complexes had been structurally resolved. Our study thus predicted high-scoring inhibitor-enzyme complexes at the plant-pathogen interface for the first time. These predicted protein complexes could be used by the community to test hypotheses about the inhibitor-enzyme structure. Second, to our knowledge, this was the first modelling of protein complexes between species using AlphaFold2 multimer. The design of AlphaFold2 multimer's 'Cross-Chain Genetics' implies the intention to construct and evaluate protein complexes within the same species (section 7.4 Evans et al. 2021). As an example, a recent large-scale *in silico* yeast-2-hybrid modelled more than 1,500 interacting pairs within *Saccharomyces cerevisiae* and detected several, previously unknown protein complexes (Humphreys et al. 2021). Inter-kingdom protein complexes at the plant-pathogen interface, such as in the tomato-*P.syringae* pv. *tomato* pathosystem, have been described experimentally (Shindo et al. 2016). The high-scoring predicted protein complexes of eukaryotic Rcr3, Pip1 and C14 in interaction with prokaryotic Cip1 are indicative that AlphaFold2 multimer could be capable of modelling inter-kingdom protein complexes. The accuracy of the predicted protein complexes remains to be confirmed experimentally.

Several open questions remain. First, predicted quality scores did not reflect differences in strength of inhibition between different targets and inhibitors, and it remains to be tested if AlphaFold2 multimer-derived scores are capable of distinguishing between protein segments of the same protein family with strong differences in target affinity. This could be tested by modelling the tomato subtilase P69B in complex with the two *Phytophthora infestans*-derived Kazal-like domains Epi1a and Epi1b, of which only the former is an inhibitor of P69B (Tian and Kamoun 2005). Furthermore, we have not tested effects of single amino acid substitutions within inhibitor-enzyme complexes. In the future, *in silico* mutating amino acids in well-characterised inhibitor-protease complexes will help

understand if AlphaFold2 multimer can accurately predict effects of single amino acid substitutions. Second, while AlphaFold2 multimer predicted high-scoring inhibitor-enzyme complexes for proteases, we do not yet understand why most glucanases and cognate inhibitors were modelled with low quality scores, despite high mean non-gap MSA depths (**Figure 49B**). Potential explanations are insufficient co-evolutionary information encoded in the amino acid MSA for *Gm*Gip1-*Ps*XEG1 and *Gm*Gip1-*Ps*XLP1, but not for *Nb*Gip2-*Pp*XEG and *Nb*Gip2-*Pp*XLP1. Additionally, for soybean endoglucanase A and *P.sojae*-derived Gip1, the stoichiometry could be different to our applied 1:1 stoichiometry. Manual modelling of the inhibitor-enzyme interaction supported a 1:1 stoichiometry (Damasceno et al. 2008), but in inhibition analyses, 1 mol of *Ps*Gip1 inhibited 3.8 mol of soybean endoglucanase A (Ham et al. 1997). For now, the low ipTM+pTM scores for endoglucanases and cognate inhibitors remains an open question. However, AlphaFold2 multimer itself offers the ability to test a multimeric soybean endoglucanase A inhibited by *Ps*Gip1.

5.4.2 Modelling over 11,000 protein pairs with AlphaFold2 multimer

We provided the ground work to utilise AlphaFold2 multimer in an unprecedented approach to identify putative, pathogen-derived hydrolase inhibitors at the plant-pathogen interface. First, we predicted small, putative apoplastic proteins for seven tomato pathogens to generate a feasible number of protein complexes to predict. Next, we modelled each of these 1,879 pathogen-derived proteins in combination with six tomato proteins resulting in 11,274 protein pairs. Last, we selected a subset of 376 preicted protein complexes for manual inspection to detect putative novel apoplastic hydrolase inhibitors.

Small, putative apoplastic proteins were predicted by a custom-made secretion pipeline which combined several layers of information previously gathered on apoplastic proteins. Modelling of the 1,879 pathogen-derived proteins in complexes with tomato proteins was achieved with a custom-made workflow of AlphaFold2 utilising high-performance computing at the University of Oxford Advanced Research Computing (ARC) facility. Predicting apoplastic proteins is, as of now, an imperfect science. Tools such as SignalP5.0 and Tar-

getP2.0 have good recall rates on test data sets at 86% and 83% respectively, but neither tool predicts protein localisation upon secretion (Armenteros et al. 2019a). We thus used ApoplastP1.0.1 which is a machine learning tool trained on experimentally validated, apoplastic proteins to infer characteristics that can be used as predictors for apoplastic localisation of secreted proteins (Sperschneider et al. 2018). While ApoplastP certainly helps to reduce the total number of secreted proteins to a subset more likely to remain in the apoplast than not, its recall rate is only 75%. Furthermore, ApoplastP had predominantly been trained on experimentally validated apoplastic proteins from fungi and oomycetes which might introduce blind spots when attempting to predict apoplastic bacterial proteins. Our additional filtering criterion of <35 kDa (including signal peptide) was another heuristic filter which will have lead to us missing novel apoplastic hydrolase inhibitors in our screen. *Nb*Gip2, as an example, has a predicted molecular weight of 47.2 kDa and would thus have been excluded from our *in silico* screen for novel apoplastic hydrolase inhibitors. Subsetting of pathogen proteins was a vital first screening step, since the generated 11,274 protein complexes already consumed 13,244.20 CPU and 8,117.84 GPU hours to predict with our custom-made AlphaFold2 multimer workflow. Generating a custom-made, more resource-efficient workflow was vital. Our pipeline improved the standard AlphaFold2 multimer workflow in two aspects by reusing outputs of database searches from AlphaFold2 monomer for each predicted complex. First, reusing data base searches from AlphaFold2 monomer resulted in only 1,885 (1,879 pathogen-derived and 6 tomato-derived proteins) time-consuming data base searches instead of 11,274. Second, reusing data base searches from AlphaFold2 monomer dramatically reduced the amount of redundant data generated. Storage capacity of only 5 TB on the ARC computing cluster proved to be a bottleneck which was resolved by using the storage predominantly for generated protein complexes, rather than for storing 11,274 overlapping data base search outputs. Open questions remain. To begin with, the selection of 1,879 pathogen-derived proteins with SignalP, TargetP and ApoplastP will have compounded prediction errors that lead to our 1,879 proteins likely missing truly apoplastic proteins. Future improvements of prediction tools should thus enhance this initial, crucial step of selecting a data

set to screen. Furthermore, our selection of small, apoplastic pathogen-derived proteins based on a molecular weight cut off will have excluded several novel hydrolase inhibitors. Modelling protein complexes with larger apoplastic, pathogen-derived proteins will likely identify additional putative protein complexes in future. In addition, since inferring accurate models of protein complexes relies on amino acid sequence information, we anticipate AlphaFold2 multimer's ability to accurately predict protein complexes to increase with growing sequencing databases. Almost half (45.02%) of the 1,879 pathogen-derived proteins showed a mean non-gap amino acid MSA depth below the saturation of information at 100. More genome sequencing, genome assemblies and structural annotations as well as better curated sequence databases will increase the ability of AlphaFold2 multimer to detect novel hydrolase inhibitors, even without increasing experimentally resolved structures. Last, it remains to be explored whether our single-variable ipTM+pTM threshold of 0.75 is the most efficient approach to select protein complex for manual screening. Given the plethora of AlphaFold2 multimer output data, a combination of more variables might result in a more efficient manual screening for hydrolase inhibitors.

5.4.3 Manual scoring of 376 protein complexes identifies novel putative apoplastic hydrolase inhibitors

We manually characterised 376 predicted protein complexes by assigning each to one of four groups which identified 23 putative hydrolase inhibitors. Most of them were annotated as hypothetical proteins and thus not previously implicated in plant immunity. Two putative hydrolase inhibitors had been described as secreted apoplastic proteins Ecp36-1 of *C.fulvum* and Six15 of *F.oxysporum* f. sp. *lycopersici*, but no biological function had been assigned to either (Mesarich et al. 2018; Simbaqueba et al. 2018). Having discovered a putative biological role for both secreted apoplastic proteins is an exciting outlook. Furthermore, having identified several putative, proteinaceous chitinase inhibitors is of paramount scientific significance. Chitinases are prominent PR proteins that have been studied extensively in plant immunity ((Punja and Zhang 1993; Grover 2012)). Despite a

large body of research on antimicrobial activity of plant chitinases and the description of variant residues in chitinases that may result from an arms race with inhbitors (Bishop et al. 2000), not a single pathogen-derived, proteinaceous chitinase inhibitor has been identified yet. While our work in this chapter has laid the foundation to apply AlphaFold2 multimer in host-pathogen interactions, several open questions emerged from our search for novel pathogen-derived hydrolase inhibitors. To begin with, we will need to confirm inhibition of targeted hydrolases experimentally. We can then expand the search for hydrolase inhibitors at the plant-pathogen interface. We expect that our selection of 1,879 pathogen-derived proteins will contain more than the 23 putative hydrolase inhibitors we identified. Other host-derived immunity-related, apoplastic enzymes such as endoglucanases, serine-like carboxypeptidases, peroxidases, oxidoreductases and others are likely to be targeted by small, pathogen-derived proteinaceous inhibitors. Exploring the 1,879 pathogen-derived proteins by studying predicted complexes with other host targets will likely yield more putative inhibitors within our screening data set. Beyond searching for novel enzyme inhibitors, our screen indicated that AlphaFold2 multimer might be able to identify protease substrates (category 2 protein pairs). As of now, identifying protease substrates is laborious and often not successful. This challenge seems to be supported by the lack of published data on protease substrates, apart from a few identified substrates such as the immunity protease cascade of P69B and Rcr3 or P69B-mediated degradation of *P.infestans* PC2 (Paulus et al. 2020; Wang et al. 2021a). To date, no substrates for immune proteases such as tomato Rcr3, C14 or Pip1 have been identified. AlphaFold2 multimer might be a novel tool to identify protease targets. Experiments in this direction could follow the conceptual approach used in this chapter and start with modelling of mature P69B with its substrate proRcr3.

5.4.4 Expression and conservation of putative hydrolase inhibitors

In these parts of the chapter, we analysed the biological context of putative hydrolase inhibitors to support their putative function at the plant-pathogen interface. We re-analysed publicly available *in planta* expression data to identify genes expressed during

infection. Furthermore, we compared the putative inhibitors to homologs in other plant pathogens to understand clade specificity and their broader evolutionary context.
Adding these two layers resulted in the identification of 13 putative hydrolase inhibitors that are expressed during infection. More *in planta* time-course RNAseq data sets could help to detect additional putative hydrolase inhibitors as we know effector expression profiles can vary during infection.

5.4.5 Novel hydrolase inhibitors predicted by AlphaFold2 multimer

All 13 putative hydrolase inhibitors are promising candidates and will be discussed based on their functional annotations. Six of the novel, putative hydrolase inhibitors are functionally annotated as hypothetical proteins. For all of them, Blastp searches against NCBI's non-redundant protein sequences did not identify any homologs with annotated functions. For these six proteins, it will be intriguing to see if future experimental characterisations will identify them as archetypes of novel apoplastic hydrolase inhibitor classes. Next, XP018243121.1 is homologous to KAF5559191.1 from *Fusarium mexicanum* which contains a 'trypsin inhibitor like cysteine rich domain' (TIL). TIL domains are described to be common among extracellular proteins and rich in disulfide bridges (see IPR002919 for a more detailed description). TIL domain-containing proteins have shown ability to inhibit trypsin-, chymotrypsin- and cathepsin-like proteases (Bania et al. 1999; Cierpicki et al. 2000; Rosengren et al. 2001). A TIL-domain containing effector of silk moth (*Bombyx mori*) showed strong inhibition of bacterial subtilisin (Li et al. 2016). Taken together, XP018243121.1, which is homologus to a trypsin inhibitor-like cysteine rich domain-containing protein, is an ideal candidate to be the first, Fol-derived inhibitor of apoplastic subtilase P69B. Six15 is another Fol-derived putative inhibitor of P69B. Six proteins have been implicated in successful host colonisation by *F.oxysporum* formae speciale and 14 Six proteins have recently been reviewed for their (a)virulence function (Jangir et al. 2021). Despite their described (a)virulence functions, no effector target of any Six protein is known (Rep et al. 2004; van der Does et al. 2008; Thatcher et al. 2012; Kashiwa et al. 2013; Gawehns et al. 2014; Ma et al. 2015; An et al. 2019). Six15 was only discovered

recently in the reference strain Fol 4278 and is described as a 'non-annotated transcript' on chromosome 14 (Simbaqueba et al. 2021). Incomplete structural annotation of Six proteins on Fol 4287 chromosome 14 had been discussed before (Schmidt et al. 2013), which is why current annotations might still not include the entire repertoire of Fol 4287 Six proteins. This explains why we had to manually add Six15 to the RefSeq annotation of Fol 4278 for this chapter. Taken together, studying Six15 in the context of colonisation and as a potential P69B inhibitor might confirm the effector target of Six15. Combining its high expression during infection and promising inhibitor complex with P69B make it a strong candidate to be the first Fusarium Six protein with an identified effector target. Next, we discovered that Ecp36 of *C.fulvum* might be a P69B inhibitor. Ecps comprise a broad range of small, secreted cystein-rich proteins by *C.fulvum* (Mesarich et al. 2018). Several Ecps are recognised by tomato plants and trigger immunity against *C.fulvum* (Laugé et al. 1998; Laugé et al. 2000). Identifying novel receptor-like proteins for novel Ecps to confer *C.fulvum* resistance in tomatoes is an ongoing field of research (Mesarich et al. 2018). In addition to characterised avirulence functions of several Ecps, Ecp1, Ecp2-1 and Ecp6 have been described for their virulence function (Laugé et al. 1997; Bolton et al. 2008; de Jonge et al. 2010; Sánchez-Vallet et al. 2013). Transcripts of *Ecp36-1* and apoplastic Ecp36-1 protein were first identified in an effectoromics screen for small, secreted *C.fulvum* proteins and nothing is known about its biological function as Ecp36-1 shows similarities to hypothetical proteins only (Mesarich et al. 2018). If Ecp36-1 proves to be a P69B inhibitor, it would be the first subtilase inhibitor described for *C.fulvum*. The importance for *C.fulvum* to prevent hydrolytic activity of P69B was highlighted by its recent implication in Avr2-Rcr3-Cf2-mediated resistance (Paulus et al. 2020). Taken together, high expression of Ecp36-1 by *C.fulvum* during infection makes Ecp36-1 a promising candidate to be the first proteinaceous inhibitor of the immuno-protease P69B of *C.fulvum*. Next, we discovered that WP046931881.1 might be a *X.perforans*-derived putative inhibitor of a class III chitinase. WP046931881.1 was annotated as a membrane-bound lysozyme inhibitor of c-type lysozymes (MliC) which are bacterial inhibitors of animal lysozymes (Callewaert et al. 2008). Given that class III chitinases have dual hydrolytic activity as chitinases and

lysozymes, detecting a bacterial protein annotated as an animal lysozyme inhibitor is not surprising (Terwisscha van Scheltinga et al. 1995). MliC proteins have so far only been found in gram-negative proteobacteria and show low sequence similarity to each other (Callewaert et al. 2008; Callewaert et al. 2012). *X.perforans*, is also a gram-negative proteobacterium. MliC proteins interact with animal lysozymes via two inhibitory loops, just as in the predicted chitinase III - WP046931881.1 protein complex (**Figure 57A**) (Callewaert et al. 2012). Despite lysozymes enhancing plant immunity against phytopathogenic bacteria, no bacterial lysozyme inhibitors of plant lysozymes have been described yet (Düring et al. 1993). Chitinase class III and its putative lysozyme activity might thus be a missing link at the plant-pathogen interface. Next, *B.cinerea* may use a Eurkayotic Phosphatidyl Ethanolamine-Binding Protein (PEBP) to inhibit P69B. PEBPs are a conserved group of proteins with a broad range of functions comprising inter- and extracelluluar proteins (Sedivy 2011). In mammals, they have been implicated in neuronal development, regulation of signalling in the MAP kinase and NF-kappaB pathways and inhibition of serine proteases [Yeung et al. 2001; Hengst et al. 2001; Corbit et al. 2003; Vallée et al. 2003 and see IPR036610]. Secreted insect PEBP has antifungal properties against a broad range of fungi (Tang et al. 2019). In plants, PEBP proteins have been implicated in seed germination, flowering and flower architecture (Karlgren et al. 2011; Yang et al. 2019). PEBPs have been implicated in a broad range of biological processes and to our knowledge, antagonistic interactions at the plant-pathogen interface in the apoplast is not one of them. It will be interesting to see if fungal PEBPs, like mammalian PEBPs, can demonstrate serine proteases inhibitory activity and suppress activity of subtilases such as P69B. Next, Fol may use Necrose-inducing secreted protein 1 (NIS1) to inhibit class III chitinase. NIS1 was first described as a small secreted protein from *Colletotrichum orbiculare* to trigger necrosis in *N.benthamiana* (Yoshino et al. 2012). Since then, homologs from several other fungi such as *Colletotrichum higginsianum, Magnaporthe oryzae* and *Valsa mali* have been studied and established NIS1 as a core fungal effector (Yoshino et al. 2012; Irieda et al. 2019; Nie et al. 2022). A C-terminal segment of NIS1 is responsible for triggering necrosis in *N.benthamiana* (Irieda et al. 2019; Nie

et al. 2022). In addition to triggering necrosis, NIS1 suppresses PAMP-triggered immunity in *N.benthamiana* (Irieda et al. 2019). NIS1 binds BAK1 and BIK1 and interferes with PR protein accumulation and ROS burst upon PAMP recognition (Irieda et al. 2019; Nie et al. 2022). It remains yet unknown how the host recognises NIS1, triggers necrosis and protects itself against NIS1 and the attempted infection. Assigning a function to chitinase class III in the context of NIS1 secretion is speculative as of now. Since *Fo*NIS1 contains part of the C-terminal sequence assumed to be involved in recognition and triggering necrosis, we hypothesise that *Fo*NIS1 is capable of triggering necrosis in tomato (Nie et al. 2022). If *Fo*NIS1 does trigger necrosis, chitinase class III could be a potential co-receptor in the fashion of Avr2-Rcr3-Cf2 (Rooney et al. 2005). Next, Fol may use cell wall protein PhiA as a putative inhibitor of a class I chitinase. PhiA is required for phialide development, asexual development and normal growth in *Aspergillus nidulans* and a PhiA homolog is transcriptionally upregulated in *A.niger* upon exposure to phenyllactic acid with antifungal properties (Melin et al. 2003; Svanström et al. 2013). Fol-derived cell wall-binding PhiA might be the first proteinaceous inhibitor of a class I chitinase which had been predicted and remained an unanswered question for over 20 years in molecular plant immunity (Bishop et al. 2000).

Taken together, this selection of 13 putative hydrolase inhibitors comprises a promising selection of proteins to experimentally test for inhibitory activity of their cognate target. For experimental characterisation, putative hydrolase inhibitors and cognate tomato proteins could be produced heterologously in *N.benthamiana* leaves via agroinfiltration or in liquid cultures of *Escherichia coli* or *Pichia pastoris*. Substrate conversion assays, activity-based protein profiling (ABPP) and co-immunoprecipitations can verify inhibitor-enzyme interactions.

Notably, nearly half of all novel, putative hydrolase inhibitors target tomato P69B. This stresses the importance of P69B in immunity of tomato, a function which we probably have not fully explored yet.

6 General discussion and conclusions

To sustainably feed a growing global population, it is important to increase plant disease resistance to reduce yield losses to plant pathogens. During plant immunity, the intercellular space (apoplast) is a crucial molecular battlefield. To understand apoplastic defence responses better, we studied biotic stress-induced secretomes of three tomato species, *N.benthamiana, G.max* and *A.thaliana*. We combined tools such as comparative proteomics of apoplastic fluids (AFs), computational analyses of molecular evolution, high-performance computing-enabled protein structure prediction and bioluminescence-based disease assays. All with the aim to identify novel proteins in apoplastic plant immunity.

6.1 Comparative mass spectrometry

In chapter 2 and 3, we analysed and compared mass spectrometry (MS) data of apoplastic fluid extracts from six different, biotic stress-induced eudicot species *Solanum lycopersicum, Solanum peruvianum, Solanum pennellii, Nicotiana benthamiana, Glycine max* and *Arabidopsis thaliana*. Biotic stress had been induced by Bion treatment, stimulating the SA signalling pathway. A first bottleneck was to extract AF with low levels of cytoplasmic contamination and while we succeeded for six species, we failed for nine. Optimising growth- and AF extraction conditions will be important to expand studies of biotic stress-induced AFs in future.

6.1.1 Generate apoplastic fluid extracts

Despite extensive experience of growing *Solanaceae* species such as *S.lycopersicum* and *N.benthamiana*, we failed to grow unstressed *Solanum pimpinellifolium, chilense* and *tuberosum*. We furthermore failed grow other eudicot species and the monocot *Hordeum vulgare* unstressed, and we could not amplify seeds for *G.soja*. Optimising growth conditions is a tedious task and parameters for future optimisation could include temperature, humidity, photoperiod and watering schedules as well as the age of plants at AF extraction

(Häffner et al. 2015; Borniego et al. 2020; Iqbal et al. 2021). Upon extraction of biotic stress-induced AFs and MS analysis, we learned that increased MDH activity is a good indicator of cytoplasmic contamination and explained differences in protein composition of *A.thaliana* samples. In future experiments, excess biological replicates of AFs should be extracted and MDH activity should be used as guidance to identify samples with low and comparable cytoplasmic contamination for accurate MS analysis of AFs.

6.1.2 Comparing Bion-induced apoplastic proteomes

To identify key similarities and differences between biotic stress-induced AFs of different species, we compared MS data based on LFQ intensities assigned to different orthogroups in chapters 2 and 3. Highly-similar was the Bion-induction and abundance of PR2 β-1,3-glucanases as well as strong induction of PR1, identifying conserved antimicrobial defence responses across the six species. In contrast, PR9 peroxidases and P69 subtilases marked stark differences between species. *Solanaceae*-specific P69 subtilases (Taylor and Qiu 2017) and their strong differences in abundance between tomato species and *N.benthamiana* raised the question of how conserved high P69 abundance is across *Solanaceae* species. Does a common recent ancestor separate high- and low P69 abundant species to indicate a single evolutionary event or has high P69 abundance upon biotic stressed been repeatedly adopted independently? To answer this question, proteomes of more *Solanaceae* species could be clustered into orthogroups to identify respective P69 homologs (Emms and Kelly 2019). Suitable *Solanaceae* species with assembled genomes or reference proteomes are *Solanum lycopersicum* (tomato) (Consortium 2012), *Solanum tuberosum* (potato) (The Potato Genome Sequencing Consortium 2011), *Solanum melongena* (aubergine) (Hirakawa et al. 2014), *Capsicum annuum* (pepper) (Kim et al. 2014b), *Petunia axillaris* (Bombarely et al. 2016) and *Petunia secreta* (Berardi et al. 2021) as well as *Nicotiana benthamiana* (wild tobacco) (Kourelis et al. 2019) and *Nicotiana tabacum* (tobacco) (Sierro et al. 2014). Next, biotic stress-induced P69 abundance in AFs could be resolved with MS analysis of Bion-induced AFs. The conceptual framework and workflow for these experiments including Bion treatment, apoplastic fluid extraction, MS sample

preparation and data analysis have been described in this thesis and could be used as guidelines.

Furthermore, we detected a noticeable disconnect between protein gel- and MS LFQ intensity of highly-abundant PR1s and PR3s (Parent et al. 1985; Jamet and Fritig 1986; Legrand et al. 1987; Joosten and De Wit 1989; Uknes et al. 1992). While LFQ intensity can be used as a proxy for protein abundance (de Godoy et al. 2008; Vogel and Marcotte 2009; Wiśniewski et al. 2012; Cox et al. 2014), their correlation remains imperfect. Biases in ionisation efficiency of peptides due to glycosylation or amino acid composition can result in disproportionately low LFQ intensity and thus an underestimation of protein abundance (Dell and Morris 2001; Peterman and Mulholland 2006; Liigand et al. 2019; Illiano et al. 2020) and computational tools are attempting to account for such biases (Dincer et al. 2022). Consistency in these biases, however, enables robust comparison between similar proteins, which is particularly accurate using MaxLFQ-normalised MS intensities, like in this thesis (Cox et al. 2014). Our data highlighted that high LFQ intensity identified highly-abundant P69s, PR2s and PR5 while low LFQ intensities of PR1s and PR3s serve as a cautionary tale to definitively associate low LFQ intensity with low protein abundance. Using LFQ intensities as proxies for protein abundance thus reliably identifies highly-abundant proteins, but will consistently underestimate the abundance of proteins with reduced ionisation efficiency. This will result in an incomplete, but not inaccurate description of highly-abundant proteins based on LFQ intensity. To understand the correlation between LFQ intensity and protein concentration of PR proteins better, AF samples could be spiked with stable isotope-labled peptides of PR proteins for comparative proteomics (Ong et al. 2002).

Comparative proteomics implicated various proteins in plant immunity: In chapter 2, we identified 39 Bion-induced orthogroups across three tomato species and in chapter 3, we identified highly-abundant SBT4 from *G.max*, PR8 from *S.peruvianum* and PR9 from *S.lycopersicum* implicating them in plant defence. To understand the full biotic stress-inducible apoplastic protein repertoire of different plant species, we proposed to analyse AFs induced by different elicitors such as flagellin, chitin, peptidoglycan and ef-

fectors (recently reviewed in Abdul Malik et al. 2020). To begin with, *S.lycopersicum* or *N.benthamiana*, both amenable to biotic stress induction and AF extraction and with extensive genomic resources available, provide ideal species for more in-depth characterisation. While chapter 2 and 3 have identified core similarities and differences between biotic stress-induced AFs of different species, we hope this to be only the starting point of understanding different apoplastic proteomic landscapes.

6.2 Molecular evolution in the era of whole genome sequencing

To further corroborate implication of tomato proteins in plant immunity, we analysed molecular evolution in tomato genes.

6.2.1 Genomics revolution and molecular evolution

In chapter 2, we took advantage of vast whole genome sequencing data to analyse molecular evolution in the form of positive selection across tomato species for genes encoding Bion-induced apoplastic tomato proteins. Therefore, we wrote a custom computational workflow to identify 21 of 51 (41.2%) of analysed genes to be under positive selection. Had we analysed positive selection on a whole-proteome basis as in Qiao et al. 2021, signal of only 5 of 51 (9.8%) genes encoding Bion-induced proteins would have been strong enough to identify them to be under positive selection due to multiple testing corrections. For now, we analysed positive selection in 12 tomato species and statistical power of selection analyses could be increased by including more publicly available genomic reads of other *Solanum* species such as wild potato and aubergine species (Aversano et al. 2015; Hardigan et al. 2017; Barchi et al. 2019; Gramazio et al. 2019; Tang et al. 2022). Furthermore, MS data from chapter 3 could be combined with our computational workflow to analyse positive selection in genes encoding Bion-induced proteins. For *A.thaliana* and *N.benthamiana* for instance, genomic sequencing data of 15 different *Nicotiana* and 13 different *Arabidopsis* species are available in NCBI's sequence read archive, as of 27.06.2022 (Leinonen et al. 2011). Genes encoding Bion-induced apoplastic proteomes could be analysed for molecular evolution following our workflow of ploidy assessment, mapping

of reads against a reference genome, variant calling, phasing of variants and positive selection analyses from recreated alleles (**chapter 2**). However, mapping reads against a single reference genome can not take advantage of the full genomic diversity of different species.

6.2.2 Pangenomics and molecular evolution

To capture the full genomic breadth of different species, pangenomics could replace single reference-based variant calling. Pangenomes generated from different closely-related species can capture loci absent from a single reference and can furthermore analyse large insertion and deletions (indels) in addition to small indels and single nucleotide polymorphisms (SNPs). No pangenome including all four tomato species clades has been generated yet. While pangenomes are becoming a more comprehensive standard in comparative genomics (Gao et al. 2019b; Sherman and Salzberg 2020; Kim et al. 2021; Qiao et al. 2021; Hoopes et al. 2022; Chapman et al. 2022), our single reference-based approach of molecular evolution already identified 21 genes encoding Bion-induced apoplastic proteins under selection.

6.3 Bioluminescent-based agromonasLux to test proteins in plant immunity

To validate genes implicated in plant immunity, we established an infection assay by incorporating bioluminescent *Pseudomonas syringae* pv *tomato* DC3000 $\Delta hopQ$ (*Pto*) into the agromonas disease assay (Buscaill et al. 2021). The assay was called agromonasLux. We established robust reduction in *Pto*-mediated bioluminescence in plants transiently expressing positive controls *Sl*FLS3 and *Nb*PR3 and compared defence-related effects at different *Pto* infection densities. Surprisingly, we found higher sensitivity to detect defence-induced reduction in luminescence at higher infection densities ($OD_{600} = 0.0010$) compared to lower infection densities ($OD_{600} = 0.0001$). After establishing agromonasLux, we tested effects of transient expression of candidate protein identified in **chapter 2**. Therefore, we cloned genes encoding 12 candidate proteins of which we only detected

three to accumulate in AF of *N.benthamiana*. Unfortunately, transient expression of none of these three proteins resulted in reduced luminescence. To improve detection of apoplastic protein accumulation, we should expand detection methods by coomassie- or ponceau-stained protein gels, substrate assays, specific antibodies or detection via MS. In the absence of detectable protein accumulation, we should improve recombinant protein expression in *N.benthamiana* to enable testing of more candidates in plant immunity. Recombinant protein expression in tobacco can be enhanced by co-expression of protease inhibitors (Grosse-Holz et al. 2018) or by expression of proteins as more stable fusion proteins such as elastin-like polypeptide fusion proteins, which increase the accumulation of recombinant proteins in tobacco while not affecting their biological activity (Patel et al. 2007; Floss et al. 2008; Conley et al. 2009). To preempt potential interference of fusion proteins with enzymatic activity, cleavable linkers could be introduced to remove the stabilising fusion protein upon purification (Chen et al. 2013). Furthermore, the assay read out of agromonasLux could be expanded to not only measure local luminescence in *Pto*-infiltrated areas, but to also assess the spread of bacteria as demonstrated by (Soldan et al. 2020). Taken together, work on agromonasLux in chapter 4 provided the foundation for rapid high-throughput disease assays to identify and characterise novel proteins in plant immunity.

6.4 Protein structure prediction to identify novel hydrolase inhibitors

In chapter 4, we took advantage of recent advances in protein structure prediction of AlphaFold2 multimer to model inhibitor-enzyme complexes at the plant-pathogen interface. We provided proof of concept that AlphaFold2 multimer identifies inhibitor-enzyme pairs based on predicted quality scores. Given sufficient sequence information, AlphaFold2 multimer predicted high-scoring complexes for 14 of 18 (77.8%) validated inhibitor-enzyme complexes, which was even higher (93.3%) for protease-inhibitor complexes only. While true positive rates were high, false positive rates were low. Of all tested non-validated interactions, 7.1% produced high-scoring predicted complexes. After excluding plausible

protein pairs, such as enzymes and cognate classes of inhibitors without validated interactions, no false positives (0.0%) were detected. We proved that AlphaFold2 multimer can be used as a high-throughput screening tool by identifying novel, putative hydrolase inhibitors. We predicted structures of >11,000 protein pairs between six apoplastic tomato hydrolases and 1,879 pathogen-derived proteins to identify 376 high-scoring predicted protein complexes. We then manually inspected all complexes, analysed publicly-available RNAseq data and compared homologs of putative inhibitors between 37 plant pathogens. The combination of AlphaFold2 multimer with these additional lines of evidence resulted in the identification of 13 novel, putative hydrolase inhibitors. Next, we will have to experimentally validate the inhibitor-hydrolase interactions. This AlphaFold2 multimer high-throughput screen was made possible by advances in machine learning, available experimentally resolved structures, growing sequencing data bases and increasing and available computing resources. These factors had moved this project into the 'adjacent possible' for biological scientists (Björneborn 2020). Growing availability of sequence databases will increase the predictive accuracy of AlphaFold2 multimer to enable fully exploring data sets such as ours, as nearly half of pathogen-derived proteins were still below ideal sequence coverage in our screen. Re-analysing our data set in future could thus reveal additional novel hydrolase inhibitors.

For plant immunity research alone, the scope of application for AlphaFold2 multimer is vast: next steps could involve using generated inhibitor-enzyme complexes for protein engineering. Furthermore, immunity signalling cascades and putative protease substrates could be studied by predicting within-species protein complexes (Humphreys et al. 2021; Tsaban et al. 2022; Bryant et al. 2022). AlphaFold2 also fuelled advances in synthetic protein design with tools such as 'AlphaDesign' (Jendrusch et al. 2021) and 'network hallucination' (Anishchenko et al. 2021), which could be used to design synthetic inhibitors for novel activity-based protein profiling (ABPP) probes. For instance, AlphaFold2 could be used to design probes that target proteases for which no ABPP probes are available to date, such as aspartic proteases. These novel probes would expand our current tool set to study immune proteases. Finally, access to AlphaFold2-mediated protein prediction is

continuously improved to lower the barrier of entry for biological scientists (Mirdita et al. 2022).

6.5 Conclusion

The aim of this thesis was to identify proteins in plant immunity. For now, we have not experimentally validated new candidate proteins in immunity. We did, however, generate and analyse large MS data sets to describe similarities and differences in plant apoplastic immunity landscapes. Furthermore, analysing molecular evolution of genes encoding Bion-induced proteins across tomato species implicated 21 proteins in immunity. Positive selection analyses identified P69B codons to test for effects in inhibitor-enzyme complexes. Moreover, we identified 13 novel, putative hydrolase inhibitors of apoplastic tomato hydrolases. In addition to generating and analysing large MS, genome sequencing and structure prediction data sets, we compiled computational analyses workflows available to the public in our GitHub repository. Last, we established a high-throughput, bioluminescence-based disease assay to test candidate proteins for their involvement in plant immunity.

References

100 Tomato Genome Sequencing Consortium (2014). "Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing". eng. In: *The Plant Journal: For Cell and Molecular Biology* 80.1, pp. 136–148. ISSN: 1365-313X. DOI: 10.1111/tpj.12616.

Abdul Malik, Nurul Azmina, Ilakiya Sharanee Kumar, and Kalaivani Nadarajah (2020). "Elicitor and receptor molecules: Orchestrators of plant defense and immunity". In: *International Journal of Molecular Sciences* 21.3, p. 963. ISSN: 1422-0067. DOI: 10.3390/ijms21030963.

Abreha, Kibrom B. et al. (2021). "Leaf apoplast of field-grown potato analyzed by quantitative proteomics and activity-based protein profiling". en. In: *International Journal of Molecular Sciences* 22.21, p. 12033. ISSN: 1422-0067. DOI: 10.3390/ijms222112033.

Acharya, Karan et al. (2013). "Overexpression of Camellia sinensis thaumatin-like protein, CsTLP in potato confers enhanced resistance to Macrophomina phaseolina and Phytophthora infestans infection". eng. In: *Molecular Biotechnology* 54.2, pp. 609–622. ISSN: 1559-0305. DOI: 10.1007/s12033-012-9603-y.

Akbudak, M. Aydın, Sukran Yildiz, and Ertugrul Filiz (2020). "Pathogenesis related protein-1 (PR-1) genes in tomato (Solanum lycopersicum L.): Bioinformatics analyses and expression profiles in response to drought stress". en. In: *Genomics* 112.6, pp. 4089–4099. ISSN: 0888-7543. DOI: 10.1016/j.ygeno.2020.07.004.

Alexander, D et al. (1993). "Increased tolerance to two oomycete pathogens in transgenic tobacco expressing pathogenesis-related protein 1a." In: *Proceedings of the National Academy of Sciences* 90.15, pp. 7327–7331. DOI: 10.1073/pnas.90.15.7327.

Ali, Ashfaq et al. (2014). "Quantitative proteomics and transcriptomics of potato in response to Phytophthora infestans in compatible and incompatible interactions". eng. In: *BMC genomics* 15, p. 497. ISSN: 1471-2164. DOI: 10.1186/1471-2164-15-497.

Almagro, L. et al. (2009). "Class III peroxidases in plant defence reactions". In: *Journal of Experimental Botany* 60.2, pp. 377–390. ISSN: 0022-0957. DOI: 10.1093/jxb/ern277.

Alvarez-Fernandez, Marcia et al. (2005). "Crystal structure of human cystatin D, a cysteine peptidase inhibitor with restricted inhibition profile". eng. In: *The Journal of Biological Chemistry* 280.18, pp. 18221–18228. ISSN: 0021-9258. DOI: 10.1074/jbc.M411914200.

An, Bang et al. (2019). "The effector SIX8 is required for virulence of Fusarium oxysporum f.sp. cubense tropical race 4 to Cavendish banana". eng. In: *Fungal Biology* 123.5, pp. 423–430. ISSN: 1878-6146. DOI: 10.1016/j.funbio.2019.03.001.

Anishchenko, Ivan et al. (2021). "De novo protein design by deep network hallucination". en. In: *Nature* 600.7889, pp. 547–552. ISSN: 1476-4687. DOI: 10.1038/s41586-021-04184-w.

Anžlovar, S. and M. Dermastia (2003). "The comparative analysis of osmotins and osmotin-like PR-5 proteins". en. In: *Plant Biology* 5.2, pp. 116–124. ISSN: 1438-8677. DOI: 10.1055/s-2003-40723.

Araki, Tomohiro and Takao Torikata (1995). "Structural classification of plant chitinases: two subclasses in class I and class II chitinase". In: *Bioscience, Biotechnology, and Biochemistry* 59.2. Publisher: Taylor & Francis _eprint: https://doi.org/10.1271/bbb.59.336, pp. 336–338. ISSN: 0916-8451. DOI: 10.1271/bbb.59.336.

Armenteros, Jose Juan Almagro et al. (2019a). "Detecting sequence signals in targeting peptides using deep learning". en. In: *Life Science Alliance* 2.5. ISSN: 2575-1077. DOI: 10.26508/lsa.201900429.

Armenteros, José Juan Almagro et al. (2019b). "SignalP 5.0 improves signal peptide predictions using deep neural networks". En. In: *Nature Biotechnology* 37.4, p. 420. ISSN: 1546-1696. DOI: 10.1038/s41587-019-0036-z.

Auwera, Geraldine A. Van der and Brian D. O'Connor (2020). *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. English. 1st edition. Sebastopol, CA: O'Reilly Media. ISBN: 978-1-4919-7519-0.

Aversano, Riccardo et al. (2015). "The Solanum commersonii genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives". In: *The Plant Cell* 27.4, pp. 954–968. ISSN: 1040-4651. DOI: 10.1105/tpc.114.135954.

Ayliffe, Michael et al. (2013). "A simple method for comparing fungal biomass in infected plant tissues". eng. In: *Molecular plant-microbe interactions: MPMI* 26.6, pp. 658–667. ISSN: 0894-0282. DOI: 10.1094/MPMI-12-12-0291-R.

Bailly, Christophe et al. (2004). "Catalase activity and expression in developing sunflower seeds as related to drying". eng. In: *Journal of Experimental Botany* 55.396, pp. 475–483. ISSN: 0022-0957. DOI: 10.1093/jxb/erh050.

Balasubramanian, Vaiyapuri et al. (2012). "Plant -1,3-glucanases: their biological functions and transgenic expression against phytopathogenic fungi". eng. In: *Biotechnology Letters* 34.11, pp. 1983–1990. ISSN: 1573-6776. DOI: 10.1007/s10529-012-1012-6.

Balint-Kurti, Peter (2019). "The plant hypersensitive response: concepts, control and consequences". In: *Molecular Plant Pathology* 20.8, pp. 1163–1178. ISSN: 1464-6722. DOI: 10.1111/mpp.12821.

Bally, Julia et al. (2018). "The rise and rise of Nicotiana benthamiana: a plant for all reasons". In: *Annual Review of Phytopathology* 56.1, pp. 405–426. DOI: 10.1146/annurev-phyto-080417-050141.

Bania, J, D Stachowiak, and A Polanowski (1999). "Primary structure and properties of the cathepsin G/chymotrypsin inhibitor from the larval hemolymph of Apis mellifera". eng. In: *European journal of biochemistry* 262.3, pp. 680–687. ISSN: 1432-1033. DOI: 10.1046/j.1432-1327.1999.00406.x.

Barchi, Lorenzo et al. (2019). "A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution". en. In: *Scientific Reports* 9.1. Number: 1 Publisher: Nature Publishing Group, p. 11769. ISSN: 2045-2322. DOI: 10.1038/s41598-019-47985-w.

Barre, Annick et al. (2000). "Purification and structural analysis of an abundant thaumatin-like protein from ripe banana fruit". en. In: *Planta* 211.6, pp. 791–799. ISSN: 1432-2048. DOI: 10.1007/s004250000354.

Bektas, Yasemin and Thomas Eulgem (2015). "Synthetic plant defense elicitors". In: *Frontiers in Plant Science* 5. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2014.00804.

Berardi, Andrea E. et al. (2021). "Complex evolution of novel red floral color in Petunia". eng. In: *The Plant Cell* 33.7, pp. 2273–2295. ISSN: 1532-298X. DOI: 10.1093/plcell/koab114.

Bhuiyan, Nazmul H et al. (2009). "Role of lignification in plant defense". In: *Plant Signaling & Behavior* 4.2, pp. 158–159. ISSN: 1559-2316. DOI: 10.4161/psb.4.2.7688.

Bishop, J. G., A. M. Dean, and T. Mitchell-Olds (2000). "Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution". In: *Proceedings of the National Academy of Sciences of the United States of America* 97.10, pp. 5322–5327. ISSN: 0027-8424. DOI: 10.1073/pnas.97.10.5322.

Bishop, J. G. et al. (2005). "Selection on Glycine beta-1,3-endoglucanase genes differentially inhibited by a Phytophthora glucanase inhibitor protein". eng. In: *Genetics* 169.2, pp. 1009–1019. ISSN: 0016-6731. DOI: 10.1534/genetics.103.025098.

Bishop, Russell E et al. (1998). "The entericidin locus of Escherichia coli and its implications for programmed bacterial cell death11Edited by M. Gottesman". en. In: *Journal of Molecular Biology* 280.4, pp. 583–596. ISSN: 0022-2836. DOI: 10.1006/jmbi.1998.1894.

Björneborn, Lennart (2020). "Adjacent Possible". en. In: *The Palgrave Encyclopedia of the Possible*. Cham: Springer International Publishing, pp. 1–12. ISBN: 9783319983905. DOI: 10.1007/978-3-319-98390-5_100-1.

Blainey, Paul, Martin Krzywinski, and Naomi Altman (2014). "Replication". en. In: *Nature Methods* 11.9, pp. 879–880. ISSN: 1548-7105. DOI: 10.1038/nmeth.3091.

Boccardo, Noelia Ayelen et al. (2019). "Expression of pathogenesis-related proteins in transplastomic tobacco plants confers resistance to filamentous pathogens under field trials". en. In: *Scientific Reports* 9.1, p. 2791. ISSN: 2045-2322. DOI: 10.1038/s41598-019-39568-6.

Bolger, Anthony et al. (2014a). "The genome of the stress-tolerant wild tomato species Solanum pennellii". en. In: *Nature Genetics* 46.9, pp. 1034–1038. ISSN: 1546-1718. DOI: 10.1038/ng.3046.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel (2014b). "Trimmomatic: a flexible trimmer for Illumina sequence data". eng. In: *Bioinformatics (Oxford, England)* 30.15, pp. 2114–2120. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btu170.

Bolton, Melvin D. et al. (2008). "The novel Cladosporium fulvum lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species". en. In: *Molecular Microbiology* 69.1, pp. 119–136. ISSN: 1365-2958. DOI: 10.1111/j.1365-2958.2008.06270.x.

Bombarely, Aureliano et al. (2016). "Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida". en. In: *Nature Plants* 2.6, pp. 1–9. ISSN: 2055-0278. DOI: 10.1038/nplants.2016.74.

Borniego, Maria L. et al. (2020). "Physiological and proteomic changes in the apoplast accompany leaf senescence in arabidopsis". In: *Frontiers in Plant Science* 10. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2019.01635.

Bowman, Shaun M. and Stephen J. Free (2006). "The structure and synthesis of the fungal cell wall". en. In: *BioEssays* 28.8, pp. 799–808. ISSN: 1521-1878. DOI: 10.1002/bies.20441.

Bozkurt, Tolga O. et al. (2011). "Phytophthora infestans effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface". en. In: *Proceedings of the National Academy of Sciences* 108.51, pp. 20832–20837. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1112708109.

Bray, Nicolas L. et al. (2016). "Near-optimal probabilistic RNA-seq quantification". en. In: *Nature Biotechnology* 34.5, pp. 525–527. ISSN: 1546-1696. DOI: 10.1038/nbt.3519.

Breen, Susan et al. (2017). "Emerging insights into the functions of pathogenesis-related protein 1". en. In: *Trends in Plant Science* 22.10, pp. 871–879. ISSN: 1360-1385. DOI: 10.1016/j.tplants.2017.06.013.

Breitenbach, Heiko H. et al. (2014). "Contrasting roles of the apoplastic aspartyl protease APOPLASTIC, ENHANCED DISEASE SUSCEPTIBILITY1-DEPENDENT1 and LEGUME LECTIN-LIKE PROTEIN1 in arabidopsis systemic acquired resistance". In: *Plant Physiology* 165.2, pp. 791–809. ISSN: 0032-0889. DOI: 10.1104/pp.114.239665.

Broglie, Karen et al. (1991). "Transgenic plants with enhanced resistance to the fungal pathogen Rhizoctonia solani". In: *Science* 254.5035. Publisher: American Association for the Advancement of Science, pp. 1194–1197. ISSN: 0036-8075. DOI: 10.1126/science.254.5035.1194.

Bryant, Patrick, Gabriele Pozzati, and Arne Elofsson (2022). "Improved prediction of protein-protein interactions using AlphaFold2". en. In: *Nature Communications* 13.1, p. 1265. ISSN: 2041-1723. DOI: 10.1038/s41467-022-28865-w.

Buell, C. Robin et al. (2003). "The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 100.18, pp. 10181–10186. ISSN: 0027-8424. DOI: 10.1073/pnas.1731982100.

Buscaill, Pierre and Renier A. L. van der Hoorn (2021). "Defeated by the nines: nine extracellular strategies to avoid microbe-associated molecular patterns recognition in plants". In: *The Plant Cell* 33.7, pp. 2116–2130. ISSN: 1040-4651. DOI: 10.1093/plcell/koab109.

Buscaill, Pierre et al. (2019). "Glycosidase and glycan polymorphism control hydrolytic release of immunogenic flagellin peptides". en. In: *Science* 364.6436, eaav0748. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aav0748.

Buscaill, Pierre et al. (2021). "Agromonas: a rapid disease assay for Pseudomonas syringae growth in agroinfiltrated leaves". en. In: *The Plant Journal* 105.3, pp. 831–840. ISSN: 1365-313X. DOI: https://doi.org/10.1111/tpj.15056.

Bushnell, Brian (2014). *BBMap: a fast, accurate, splice-aware aligner*. English. Tech. rep. LBNL-7065E. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

Cai, Kewei et al. (2021). "Genome-wide identification and analysis of class III peroxidases in Betula pendula". In: *BMC Genomics* 22.1, p. 314. ISSN: 1471-2164. DOI: 10.1186/s12864-021-07622-1.

Caicedo, Ana Lucia, Barbara A. Schaal, and Barbara N. Kunkel (1999). "Diversity and molecular evolution of the RPS2 resistance gene in Arabidopsis thaliana". In: *Proceedings of the National Academy of Sciences* 96.1, pp. 302–306. DOI: 10.1073/pnas.96.1.302.

Callewaert, Lien et al. (2008). "A new family of lysozyme inhibitors contributing to lysozyme tolerance in gram-negative bacteria". In: *PLoS Pathogens* 4.3, e1000019. ISSN: 1553-7366. DOI: 10.1371/journal.ppat.1000019.

Callewaert, Lien et al. (2012). "Guards of the great wall: bacterial lysozyme inhibitors". en. In: *Trends in Microbiology* 20.10, pp. 501–510. ISSN: 0966-842X. DOI: 10.1016/j.tim.2012.06.005.

Cao, H. et al. (1994). "Characterization of an Arabidopsis mutant that Is nonresponsive to inducers of systemic acquired resistance." In: *The Plant Cell* 6.11, pp. 1583–1592. ISSN: 1040-4651. DOI: 10.1105/tpc.6.11.1583.

Cao, Yunpeng et al. (2016). "Structural, evolutionary, and functional analysis of the class III peroxidase gene family in chinese pear (Pyrus bretschneideri)". In: *Frontiers in Plant Science* 7. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2016.01874.

Chandrasekar, Balakumaran, Tram Ngoc Hong, and Renier A. L. van der Hoorn (2017). "Inhibitor discovery by convolution ABPP". eng. In: *Methods in Molecular Biology (Clifton, N.J.)* 1491, pp. 47–56. ISSN: 1940-6029. DOI: 10.1007/978-1-4939-6439-0_4.

Chapman, Mark A., Yuqi He, and Meiliang Zhou (2022). "Beyond a reference genome: pangenomes and population genomics of underutilized and orphan crops for future food and nutrition security". en. In: *New Phytologist* 234.5, pp. 1583–1597. ISSN: 1469-8137. DOI: 10.1111/nph.18021.

Chen, S. X. and P. Schopfer (1999). "Hydroxyl-radical production in physiological reactions. A novel function of peroxidase". eng. In: *European Journal of Biochemistry* 260.3, pp. 726–735. ISSN: 0014-2956. DOI: 10.1046/j.1432-1327.1999.00199.x.

Chen, Xiaoying, Jennica Zaro, and Wei-Chiang Shen (2013). "Fusion protein linkers: Property, design and functionality". In: *Advanced drug delivery reviews* 65.10, pp. 1357–1369. ISSN: 0169-409X. DOI: 10.1016/j.addr.2012.09.039.

Chen, Ying-Lan et al. (2014). "Quantitative peptidomics study reveals that a wound-induced peptide from PR-1 regulates immune signaling in tomato". eng. In: *The Plant Cell* 26.10, pp. 4135–4148. ISSN: 1532-298X. DOI: 10.1105/tpc.114.131185.

Choi, Hyong Woo et al. (2007). "Hydrogen peroxide generation by the pepper extracellular peroxidase CaPO2 activates local and systemic cell death and defense response to bacterial pathogens". eng. In: *Plant Physiology* 145.3, pp. 890–904. ISSN: 0032-0889. DOI: 10.1104/pp.107.103325.

Chu, Ming-Hung et al. (2011). "Crystal structure of tarocystatin–papain complex: implications for the inhibition property of group-2 phytocystatins". In: *Planta* 234.2, pp. 243–254. ISSN: 0032-0935. DOI: 10.1007/s00425-011-1398-8.

Cierpicki, T, J Bania, and J Otlewski (2000). "NMR solution structure of Apis mellifera chymotrypsin/cathepsin G inhibitor-1 (AMCI-1): structural similarity with Ascaris protease inhibitors". eng. In: *Protein science* 9.5, pp. 976–984. ISSN: 1469-896X. DOI: 10.1110/ps.9.5.976.

Clark, Kelley et al. (2018). "An effector from the Huanglongbing-associated pathogen targets citrus proteases". en. In: *Nature Communications* 9.1, pp. 1–11. ISSN: 2041-1723. DOI: 10.1038/s41467-018-04140-9.

Cock, Peter J. A. et al. (2009). "Biopython: freely available Python tools for computational molecular biology and bioinformatics". In: *Bioinformatics* 25.11, pp. 1422–1423. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btp163.

Collinge, David B. et al. (1993). "Plant chitinases". en. In: *The Plant Journal* 3.1, pp. 31–40. ISSN: 1365-313X. DOI: 10.1046/j.1365-313X.1993.t01-1-00999.x.

Conley, Andrew J. et al. (2009). "Optimization of elastin-like polypeptide fusions for expression and purification of recombinant proteins in plants". en. In: *Biotechnology and Bioengineering* 103.3, pp. 562–573. ISSN: 1097-0290. DOI: 10.1002/bit.22278.

Consortium, The Tomato Genome (2012). "The tomato genome sequence provides insights into fleshy fruit evolution". en. In: *Nature* 485.7400, pp. 635–641. ISSN: 1476-4687. DOI: 10.1038/nature11119.

Corbit, Kevin C. et al. (2003). "Activation of Raf-1 signaling by protein kinase C through a mechanism involving Raf kinase inhibitory protein". English. In: *Journal of Biological Chemistry* 278.15, pp. 13061–13068. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M210015200.

Cornelissen, Ben J. C., Rob A. M. Hooft van Huijsduijnen, and John F. Bol (1986). "A tobacco mosaic virus-induced tobacco protein is homologous to the sweet-tasting protein thaumatin". en. In: *Nature* 321.6069, pp. 531–532. ISSN: 1476-4687. DOI: 10.1038/321531a0.

Corwin, Jason A. and Daniel J. Kliebenstein (2017). "Quantitative resistance: more than just perception of a pathogen". In: *The Plant Cell* 29.4, pp. 655–665. ISSN: 1040-4651. DOI: 10.1105/tpc.16.00915.

Cox, Jürgen and Matthias Mann (2008). "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quan-

tification". eng. In: *Nature Biotechnology* 26.12, pp. 1367–1372. ISSN: 1546-1696. DOI: 10.1038/nbt.1511.

Cox, Jürgen et al. (2011). "Andromeda: a peptide search engine integrated into the MaxQuant environment". eng. In: *Journal of Proteome Research* 10.4, pp. 1794–1805. ISSN: 1535-3907. DOI: 10.1021/pr101065j.

Cox, Jürgen et al. (2014). "Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ". In: *Molecular & Cellular Proteomics : MCP* 13.9, pp. 2513–2526. ISSN: 1535-9476. DOI: 10.1074/mcp.M113.031591.

Crété, Rémi et al. (2020). "Rotating and stacking genes can improve crop resistance durability while potentially selecting highly virulent pathogen strains". en. In: *Scientific Reports* 10.1, p. 19752. ISSN: 2045-2322. DOI: 10.1038/s41598-020-76788-7.

Cutt, John R. et al. (1989). "Disease response to tobacco mosaic virus in transgenic tobacco plants that constitutively express the pathogenesis-related PR1b gene". en. In: *Virology* 173.1, pp. 89–97. ISSN: 0042-6822. DOI: 10.1016/0042-6822(89)90224-9.

Damasceno, Cynthia M. B. et al. (2008). "Structure of the glucanase inhibitor protein (GIP) family from phytophthora species suggests coevolution with plant endo-beta-1,3-glucanases". eng. In: *Molecular plant-microbe interactions: MPMI* 21.6, pp. 820–830. ISSN: 0894-0282. DOI: 10.1094/MPMI-21-6-0820.

Datta, Swapan K. and Subbaratnam Muthukrishnan (1999). "Functions and regulation of plant -1,3-glucanases (PR-2)". In: *Pathogenesis-Related Proteins in Plants*. CRC Press. ISBN: 978-0-429-11508-0.

Daudi, Arsalan et al. (2012). "The apoplastic oxidative burst peroxidase in Arabidopsis is a major component of pattern-triggered immunity". eng. In: *The Plant Cell* 24.1, pp. 275–287. ISSN: 1532-298X. DOI: 10.1105/tpc.111.093039.

Dawkins, Richard and John Richard Krebs (1979). "Arms races between and within species". en. In: *Proc. R. Soc. Lond. B* 205.1161, pp. 489–511. ISSN: 0080-4649, 2053-9193. DOI: 10.1098/rspb.1979.0081.

De Godoy, Lyris M. F. et al. (2008). "Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast". eng. In: *Nature* 455.7217, pp. 1251–1254. ISSN: 1476-4687. DOI: 10.1038/nature07341.

De Jonge, Ronnie et al. (2010). "Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants". In: *Science* 329.5994, pp. 953–955. DOI: 10.1126/science.1190859.

Dean, Ralph et al. (2012). "The Top 10 fungal pathogens in molecular plant pathology". en. In: *Molecular Plant Pathology* 13.4, pp. 414–430. ISSN: 1364-3703. DOI: 10.1111/j.1364-3703.2011.00783.x.

Delaney, T P, L Friedrich, and J A Ryals (1995). "Arabidopsis signal transduction mutant defective in chemically and biologically induced disease resistance." In: *Proceedings of the National Academy of Sciences* 92.14, pp. 6602–6606. DOI: 10.1073/pnas.92.14.6602.

Delaney, T. P. et al. (1994). "A central role of salicylic acid in plant disease resistance". eng. In: *Science (New York, N.Y.)* 266.5188, pp. 1247–1250. ISSN: 0036-8075. DOI: 10.1126/science.266.5188.1247.

Dell, Anne and Howard R. Morris (2001). "Glycoprotein structure determination by mass spectrometry". In: *Science* 291.5512, pp. 2351–2356. DOI: 10.1126/science.1058890.

Dempsey, D'Maris Amick and Daniel F. Klessig (2017). "How does the multifaceted plant hormone salicylic acid combat disease in plants and are similar mechanisms utilized in humans?" In: *BMC Biology* 15.1, p. 23. ISSN: 1741-7007. DOI: 10.1186/s12915-017-0364-8.

Derache, Chrystelle et al. (2012). "Crystal structure of greglin, a novel non-classical Kazal inhibitor, in complex with subtilisin". en. In: *The FEBS Journal* 279.24, pp. 4466–4478. ISSN: 1742-4658. DOI: 10.1111/febs.12033.

Dincer, Ayse et al. (2022). *Reducing peptide sequence bias in quantitative mass spectrometry data with machine learning*. en. DOI: 10.1101/2022.04.11.487945.

Ding, Yuli et al. (2018). "Opposite roles of salicylic acid receptors NPR1 and NPR3/NPR4 in transcriptional regulation of plant immunity". en. In: *Cell* 173.6, 1454–1467.e15. ISSN: 0092-8674. DOI: 10.1016/j.cell.2018.03.044.

Doehlemann, Gunther and Christoph Hemetsberger (2013). "Apoplastic immunity and its suppression by filamentous plant pathogens". eng. In: *The New Phytologist* 198.4, pp. 1001–1016. ISSN: 1469-8137. DOI: 10.1111/nph.12277.

Doehlemann, Gunther et al. (2009). "Pep1, a secreted effector protein of Ustilago maydis, is required for successful invasion of plant cells". eng. In: *PLoS pathogens* 5.2, e1000290. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1000290.

Doxey, Andrew C. et al. (2007). "Functional divergence in the Arabidopsis beta-1,3-glucanase gene family inferred by phylogenetic reconstruction of expression states". eng. In: *Molecular Biology and Evolution* 24.4, pp. 1045–1055. ISSN: 0737-4038. DOI: 10.1093/molbev/msm024.

Düring, Klaus et al. (1993). "Transgenic potato plants resistant to the phytopathogenic bacterium Erwinia carotovora". en. In: *The Plant Journal* 3.4, pp. 587–598. ISSN: 1365-313X. DOI: 10.1046/j.1365-313X.1993.03040587.x.

Elias, Joshua E. and Steven P. Gygi (2007). "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry". en. In: *Nature Methods* 4.3, pp. 207–214. ISSN: 1548-7105. DOI: 10.1038/nmeth1019.

Emms, D. M. and S. Kelly (2018). "STAG: Species Tree Inference from All Genes". en. In: Type: article, p. 267914. DOI: https://doi.org/10.1101/267914.

Emms, David M. and Steven Kelly (2017). "STRIDE: Species Tree Root Inference from Gene Duplication Events". eng. In: *Molecular Biology and Evolution* 34.12, pp. 3267–3278. ISSN: 1537-1719. DOI: 10.1093/molbev/msx259.

— (2019). "OrthoFinder: Phylogenetic orthology inference for comparative genomics". In: *Genome Biology* 20.1, p. 238. ISSN: 1474-760X. DOI: 10.1186/s13059-019-1832-y.

Engler, Carola et al. (2014). "A golden gate modular cloning toolbox for plants". eng. In: *ACS synthetic biology* 3.11, pp. 839–843. ISSN: 2161-5063. DOI: 10.1021/sb4001504.

Evans, Richard et al. (2021). "Protein complex prediction with AlphaFold-Multimer". en. In: Type: article, p. 2021.10.04.463034. DOI: https://doi.org/10.1101/2021.10.04.463034.

Fan, Jun, Casey Crooks, and Chris Lamb (2008). "High-throughput quantitative luminescence assay of the growth in planta of Pseudomonas syringae chromosomally tagged with Photorhabdus luminescens luxCDABE". eng. In: *The Plant Journal: For Cell and Molecular Biology* 53.2, pp. 393–399. ISSN: 0960-7412. DOI: 10.1111/j.1365-313X.2007.03303.x.

Fesel, Philipp H. and Alga Zuccaro (2016). "-glucan: Crucial component of the fungal cell wall and elusive MAMP in plants". en. In: *Fungal Genetics and Biology* 90, pp. 53–60. ISSN: 1087-1845. DOI: 10.1016/j.fgb.2015.12.004.

Fierens, Ellen et al. (2007). "TLXI, a novel type of xylanase inhibitor from wheat (Triticum aestivum) belonging to the thaumatin family". In: *Biochemical Journal* 403.Pt 3, pp. 583–591. ISSN: 0264-6021. DOI: 10.1042/BJ20061291.

Finkers, Richard et al. (2008). "Quantitative resistance to Botrytis cinerea from Solanum neorickii". en. In: *Euphytica* 159.1, pp. 83–92. ISSN: 1573-5060. DOI: 10.1007/s10681-007-9460-0.

Flach, J., P. E. Pilet, and P. Jollès (1992). "What's new in chitinase research?" eng. In: *Experientia* 48.8, pp. 701–716. ISSN: 0014-4754. DOI: 10.1007/BF02124285.

Floss, Doreen M. et al. (2008). "Biochemical and functional characterization of anti-HIV antibody-ELP fusion proteins from transgenic plants". eng. In: *Plant Biotechnology Journal* 6.4, pp. 379–391. ISSN: 1467-7652. DOI: 10.1111/j.1467-7652.2008.00326.x.

Gaffney, Thomas et al. (1993). "Requirement of salicylic acid for the induction of systemic acquired resistance". en. In: *Science* 261.5122, pp. 754–756. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.261.5122.754.

Gamir, Jordi et al. (2017). "The sterol-binding activity of pathogenesis-related protein 1 reveals the mode of action of an antimicrobial protein". en. In: *The Plant Journal* 89.3, pp. 502–509. ISSN: 1365-313X. DOI: 10.1111/tpj.13398.

Gao, Feng et al. (2019a). "Deacetylation of chitin oligomers increases virulence in soil-borne fungal pathogens". eng. In: *Nature Plants* 5.11, pp. 1167–1176. ISSN: 2055-0278. DOI: 10.1038/s41477-019-0527-4.

Gao, Lei et al. (2019b). "The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor". En. In: *Nature Genetics*, p. 1. ISSN: 1546-1718. DOI: 10.1038/s41588-019-0410-2.

Gao, Qing-Ming et al. (2015). "Signal regulators of systemic acquired resistance". In: *Frontiers in Plant Science* 6. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2015.00228.

Gawehns, F. et al. (2014). "The Fusarium oxysporum effector Six6 contributes to virulence and suppresses I-2-mediated cell death". eng. In: *Molecular plant-microbe interactions: MPMI* 27.4, pp. 336–348. ISSN: 0894-0282. DOI: 10.1094/MPMI-11-13-0330-R.

Geneious, Prime (2022). *Geneious Prime*.

Ghosh, Raka and Chandana Chakrabarti (2008). "Crystal structure analysis of NP24-I: a thaumatin-like protein". en. In: *Planta* 228.5, p. 883. ISSN: 1432-2048. DOI: 10.1007/s00425-008-0790-5.

Gianinazzi, S. and B.YR 1974 Kassanis (1974). "Virus resistance induced in plants by polyacrylic acid". In: *Journal of General Virology* 23.1, pp. 1–9. ISSN: 1465-2099, DOI: 10.1099/0022-1317-23-1-1.

Glazebrook, J., E. E. Rogers, and F. M. Ausubel (1996). "Isolation of Arabidopsis mutants with enhanced disease susceptibility by direct screening". eng. In: *Genetics* 143.2, pp. 973–982. ISSN: 0016-6731. DOI: 10.1093/genetics/143.2.973.

Glazebrook, Jane (2005). "Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens". eng. In: *Annual Review of Phytopathology* 43, pp. 205–227. ISSN: 0066-4286. DOI: 10.1146/annurev.phyto.43.040204.135923.

Godson, Alice and Renier A L van der Hoorn (2021). "The front line of defence: a meta-analysis of apoplastic proteases in plant immunity". In: *Journal of Experimental Botany* 72.9, pp. 3381–3394. ISSN: 0022-0957. DOI: 10.1093/jxb/eraa602.

Goldstein, I. J., C. E. Hollerman, and E. E. Smith (1965). "Protein-carbohydrate interation. II Inhibition studies on the interaction of Concanavalin A with polysaccharides." eng. In: *Biochemistry* 4, pp. 876–883. ISSN: 0006-2960. DOI: 10.1021/bi00881a013.

Graham, M. Y et al. (2003). "Induced expression of pathogenesis-related protein genes in soybean by wounding and the Phytophthora sojae cell wall glucan elicitor". en. In: *Physiological and Molecular Plant Pathology* 63.3, pp. 141–149. ISSN: 0885-5765. DOI: 10.1016/j.pmpp.2003.11.002.

Gramazio, Pietro et al. (2019). "Whole-Genome resequencing of seven eggplant (Solanum melongena) and one wild relative (S. incanum) accessions provides new insights and breeding tools for eggplant enhancement". English. In: *Frontiers in Plant Science* 10. ISSN: 1664-462X. DOI: 10.3389/fpls.2019.01220.

Grenier, Jean et al. (1999). "Some thaumatin-like proteins hydrolyse polymeric -1,3-glucans". en. In: *The Plant Journal* 19.4, pp. 473–480. ISSN: 1365-313X. DOI: 10.1046/j.1365-313X.1999.00551.x.

Grison, R. et al. (1996). "Field tolerance to fungal pathogens of Brassica napus constitutively expressing a chimeric chitinase gene". eng. In: *Nature Biotechnology* 14.5, pp. 643–646. ISSN: 1087-0156. DOI: 10.1038/nbt0596-643.

Grosse-Holz, Friederike et al. (2018). "Three unrelated protease inhibitors enhance accumulation of pharmaceutical recombinant proteins in Nicotiana benthamiana". en. In: *Plant Biotechnology Journal* 16.10, pp. 1797–1810. ISSN: 1467-7652. DOI: 10.1111/pbi.12916.

Grosse-Holz, Friederike et al. (2018). "The transcriptome, extracellular proteome and active secretome of agroinfiltrated Nicotiana benthamiana uncover a large, diverse protease repertoire". In: *Plant Biotechnology Journal* 16.5, pp. 1068–1084. ISSN: 1467-7644. DOI: 10.1111/pbi.12852.

Grover, Anita (2012). "Plant chitinases: Genetic diversity and physiological roles". In: *Critical Reviews in Plant Sciences* 31.1, pp. 57–73. ISSN: 0735-2689. DOI: 10.1080/07352689.2011.616043.

Guo, Baodian et al. (2019). "Phytophthora sojae effector PsAvh240 inhibits host aspartic protease secretion to promote infection". en. In: *Molecular Plant* 12.4, pp. 552–564. ISSN: 1674-2052. DOI: 10.1016/j.molp.2019.01.017.

Haas, Brian J. et al. (2009). "Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans". en. In: *Nature* 461.7262, pp. 393–398. ISSN: 1476-4687. DOI: 10.1038/nature08358.

Ham, Kyung-Sik et al. (1997). "Fungal pathogens secrete an inhibitor protein that distinguishes isoforms of plant pathogenesis-related endo--1,3-glucanases". en. In: *The Plant Journal* 11.2, pp. 169–179. ISSN: 1365-313X. DOI: 10.1046/j.1365-313X.1997.11020169.x.

Hamamouch, Noureddine et al. (2012). "The interaction of the novel 30C02 cyst nematode effector protein with a plant -1,3-endoglucanase may suppress host defence to promote parasitism". In: *Journal of Experimental Botany* 63.10, pp. 3683–3695. ISSN: 0022-0957. DOI: 10.1093/jxb/ers058.

Han, Li-Bo et al. (2019). "The cotton apoplastic protein CRR1 stabilizes chitinase 28 to facilitate defense against the fungal pathogen Verticillium dahliae". eng. In: *The Plant Cell* 31.2, pp. 520–536. ISSN: 1532-298X. DOI: 10.1105/tpc.18.00390.

Han, Sang-Wook et al. (2011). "Small protein-mediated quorum sensing in a gram-negative bacterium". en. In: *PLOS ONE* 6.12, e29192. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0029192.

Hardigan, Michael A. et al. (2017). "Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato". In: *Proceedings of the National Academy of Sciences* 114.46. Publisher: Proceedings of the National Academy of Sciences, E9999–E10008. DOI: 10.1073/pnas.1714380114.

Helleboid, Stéphane et al. (1998). "Extracellular -1,3-glucanases are induced during early somatic embryogenesis in Cichorium". en. In: *Planta* 205.1, pp. 56–63. ISSN: 1432-2048. DOI: 10.1007/s004250050296.

Hemetsberger, Christoph et al. (2012). "The Ustilago maydis effector Pep1 suppresses plant immunity by inhibition of host peroxidase activity". In: *PLoS Pathogens* 8.5. ISSN: 1553-7366. DOI: 10.1371/journal.ppat.1002684.

Hengst, Ulrich et al. (2001). "The phosphatidylethanolamine-binding protein Is the prototype of a novel family of serine protease inhibitors". en. In: *Journal of Biological Chemistry* 276.1, pp. 535–540. ISSN: 0021-9258. DOI: 10.1074/jbc.M002524200.

Hind, Sarah R. et al. (2016). "Tomato receptor FLAGELLIN-SENSING 3 binds flgII-28 and activates the plant immune system". eng. In: *Nature Plants* 2, p. 16128. ISSN: 2055-0278. DOI: 10.1038/nplants.2016.128.

Hirakawa, Hideki et al. (2014). "Draft genome sequence of eggplant (Solanum melongena L.): the representative solanum species indigenous to the old world". eng. In: *DNA research: an international journal for rapid publication of reports on genes and genomes* 21.6, pp. 649–660. ISSN: 1756-1663. DOI: 10.1093/dnares/dsu027.

Holcomb, Joshua et al. (2017). "Protein crystallization: Eluding the bottleneck of X-ray crystallography". eng. In: *AIMS biophysics* 4.4, pp. 557–575. ISSN: 2377-9098. DOI: 10.3934/biophy.2017.4.557.

Homma, Felix (2022a). "A secretion prediction pipeline". In: *Zenodo*. DOI: 10.5281/ZENODO.7424834.

— (2022b). "An AlphaFold2 multimer workflow to identify high-scoring protein complexes". In: *Zenodo*. DOI: 10.5281/ZENODO.7424746.

— (2022c). "CDS extraction pipeline". In: *Zenodo*. DOI: 10.5281/ZENODO.7424845.

— (2022d). "Identifying apoplastic tomato proteins in immunity by combining comparative proteomics with molecular evolution". In: *Zenodo*. DOI: 10.5281/ZENODO.7424842.

— (2022e). "Phasing pipeline". In: *Zenodo*. DOI: 10.5281/ZENODO.7424853.

— (2022f). "Positive selection pipeline". In: *Zenodo*. DOI: 10.5281/ZENODO.7424856.

— (2022g). "Variant calling pipeline". In: *Zenodo*. DOI: 10.5281/ZENODO.7424860.

Hoopes, Genevieve et al. (2022). "Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity". en. In: *Molecular Plant* 15.3, pp. 520–536. ISSN: 1674-2052. DOI: 10.1016/j.molp.2022.01.003.

Huang, Jie et al. (2020). "Phytophthora effectors modulate genome-wide alternative splicing of host mRNAs to reprogram plant immunity". en. In: *Molecular Plant*. Special Issue on Plant-Microbe Interactions 13.10, pp. 1470–1484. ISSN: 1674-2052. DOI: 10.1016/j.molp.2020.07.007.

Huang, Xu-Xu et al. (2021). "The novel pathogen-responsive glycosyltransferase UGT73C7 mediates the redirection of phenylpropanoid metabolism and promotes SNC1-dependent Arabidopsis immunity". en. In: *The Plant Journal* 107.1, pp. 149–165. ISSN: 1365-313X. DOI: 10.1111/tpj.15280.

Humphreys, Ian R. et al. (2021). "Computed structures of core eukaryotic protein complexes". In: *Science* 374.6573, eabm4805. DOI: 10.1126/science.abm4805.

Hurlbert, Stuart H. (1984). "Pseudoreplication and the design of ecological field experiments". en. In: *Ecological Monographs* 54.2, pp. 187–211. ISSN: 1557-7015. DOI: 10.2307/1942661.

Häffner, Eva, Sandra Konietzki, and Elke Diederichsen (2015). "Keeping control: the role of senescence and development in plant pathogenesis and defense". In: *Plants* 4.3, pp. 449–488. ISSN: 2223-7747. DOI: 10.3390/plants4030449.

Hörger, Anja C and Renier A. L. van der Hoorn (2013). "The structural basis of specific protease–inhibitor interactions at the plant–pathogen interface". In: *Current Opinion in Structural Biology*. Catalysis and regulation / Protein-protein interactions 23.6, pp. 842–850. ISSN: 0959-440X. DOI: 10.1016/j.sbi.2013.07.013.

Ibeas, José I. et al. (2000). "Fungal cell wall phosphomannans facilitate the toxic activity of a plant PR-5 protein". en. In: *The Plant Journal* 23.3, pp. 375–383. ISSN: 1365-313X. DOI: 10.1046/j.1365-313x.2000.00792.x.

Iida, Yuichiro et al. (2015). "Novel mutations detected in avirulence genes overcoming tomato Cf resistance genes in isolates of a Japanese population of Cladosporium fulvum". eng. In: *PloS One* 10.4, e0123271. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0123271.

Illiano, Anna et al. (2020). "Protein glycosylation investigated by mass spectrometry: an overview". In: *Cells* 9.9, p. 1986. ISSN: 2073-4409. DOI: 10.3390/cells9091986.

Ilyas, Muhammad et al. (2015). "Functional divergence of two secreted immune proteases of tomato". In: *Current Biology* 25.17, pp. 2300–2306. ISSN: 0960-9822. DOI: 10.1016/j.cub.2015.07.030.

Iqbal, Zahra et al. (2021). "Plant defense responses to biotic stress and Its interplay with fluctuating dark/light conditions". In: *Frontiers in Plant Science* 12. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2021.631810.

Irieda, Hiroki et al. (2019). "Conserved fungal effector suppresses PAMP-triggered immunity by targeting plant immune kinases". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 116.2, pp. 496–505. ISSN: 1091-6490. DOI: 10.1073/pnas.1807297116.

Jamet, E. and B. Fritig (1986). "Purification and characterization of 8 of the pathogenesis-related proteins in tobacco leaves reacting hypersensitively to tobacco mosaic virus". eng. In: *Plant Molecular Biology* 6.2, pp. 69–80. ISSN: 0167-4412. DOI: 10.1007/BF00027300.

Jangir, Pooja et al. (2021). "Secreted in xylem genes: Drivers of host adaptation in Fusarium oxysporum". In: *Frontiers in Plant Science* 12, p. 628611. ISSN: 1664-462X. DOI: 10.3389/fpls.2021.628611.

Jashni, Mansoor Karimi et al. (2015). "Synergistic action of a metalloprotease and a serine protease from Fusarium oxysporum f. sp. lycopersici cleaves chitin-binding tomato chitinases, reduces their antifungal activity, and enhances fungal virulence". eng. In: *Molecular plant-microbe interactions: MPMI* 28.9, pp. 996–1008. ISSN: 0894-0282. DOI: 10.1094/MPMI-04-15-0074-R.

Jendrusch, Michael, Jan O. Korbel, and S. Kashif Sadiq (2021). *AlphaDesign: A de novo protein design framework based on AlphaFold*. en. DOI: 10.1101/2021.10.11.463937.

Jones, Jonathan D. G. and Jeffery L. Dangl (2006). "The plant immune system". en. In: *Nature* 444.7117, pp. 323–329. ISSN: 1476-4687. DOI: 10.1038/nature05286.

Jones, Philip et al. (2014). "InterProScan 5: genome-scale protein function classification". en. In: *Bioinformatics* 30.9, pp. 1236–1240. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btu031.

Joosten, Matthieu H. A. J. and Pierre J. G. M. De Wit (1989). "Identification of several pathogenesis-related proteins in tomato leaves inoculated with Cladosporium fulvum (syn. Fulvia fulva) as 1,3--glucanases and chitinases 1". In: *Plant Physiology* 89.3, pp. 945–951. ISSN: 0032-0889. DOI: 10.1104/pp.89.3.945.

Jumper, John et al. (2021). "Highly accurate protein structure prediction with AlphaFold". en. In: *Nature* 596.7873, pp. 583–589. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03819-2.

Jutras, Philippe V. et al. (2021). "AgroLux: bioluminescent Agrobacterium to improve molecular pharming and study plant immunity". en. In: *The Plant Journal* n/a.n/a. ISSN: 1365-313X. DOI: 10.1111/tpj.15454.

Kahlon, Parvinderdeep S. et al. (2020). "Population studies of the wild tomato species Solanum chilense reveal geographically structured major gene-mediated pathogen resistance". en. In: *bioRxiv*, p. 2020.05.29.122960. DOI: 10.1101/2020.05.29.122960.

Kahlon, Parvinderdeep S. et al. (2021). "Quantitative resistance differences between and within natural populations of Solanum chilense against the oomycete pathogen Phytophthora infestans". eng. In: *Ecology and Evolution* 11.12, pp. 7768–7778. ISSN: 2045-7758. DOI: 10.1002/ece3.7610.

Kamoun, Sophien (2022). "Benthamiana rhapsody". In: DOI: https://doi.org/10.5281/zenodo.5866234.

Kamoun, Sophien et al. (2015). "The Top 10 oomycete pathogens in molecular plant pathology". en. In: *Molecular Plant Pathology* 16.4, pp. 413–434. ISSN: 1364-3703. DOI: 10.1111/mpp.12190.

Karlgren, Anna et al. (2011). "Evolution of the PEBP gene family in plants: Functional diversification in seed plant evolution". In: *Plant Physiology* 156.4, pp. 1967–1977. ISSN: 0032-0889. DOI: 10.1104/pp.111.176206.

Kaschani, Farnusch and Renier A. L. van der Hoorn (2011). "A model of the C14-EPIC complex indicates hotspots for a protease-inhibitor arms race in the oomycete-potato interaction". In: *Plant Signaling & Behavior* 6.1, pp. 109–112. ISSN: 1559-2316. DOI: 10.4161/psb.6.1.14190.

Kaschani, Farnusch et al. (2010). "An effector-targeted protease contributes to defense against Phytophthora infestans and is under diversifying selection in natural hosts1[W]". In: *Plant Physiology* 154.4, pp. 1794–1804. ISSN: 0032-0889. DOI: 10.1104/pp.110.158030.

Kashiwa, Takeshi et al. (2013). "An avirulence gene homologue in the tomato wilt fungus Fusarium oxysporum f. sp. lycopersici race 1 functions as a virulence gene in the cabbage yellows fungus F. oxysporum f. sp. conglutinans". en. In: *Journal of General Plant Pathology* 79.6, pp. 412–421. ISSN: 1610-739X. DOI: 10.1007/s10327-013-0471-5.

Katoh, Kazutaka and Daron M. Standley (2013). "MAFFT multiple sequence slignment software version 7: Improvements in performance and usability". en. In: *Molecular Biology and Evolution* 30.4, pp. 772–780. ISSN: 0737-4038. DOI: 10.1093/molbev/mst010.

Katz, Lee S. et al. (2019). "Mashtree: a rapid comparison of whole genome sequence files". en. In: *Journal of Open Source Software* 4.44, p. 1762. ISSN: 2475-9066. DOI: 10.21105/joss.01762.

Keller, Oliver et al. (2008). "Scipio: Using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species". In: *BMC Bioinformatics* 9.1, p. 278. ISSN: 1471-2105. DOI: 10.1186/1471-2105-9-278.

Kelly, Steven and Philip K. Maini (2013). "DendroBLAST: approximate phylogenetic trees in the absence of multiple sequence alignments". eng. In: *PloS One* 8.3, e58537. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0058537.

Khokhani, Devanshi et al. (2017). "A single regulator mediates strategic switching between attachment/spread and growth/virulence in the plant pathogen Ralstonia solanacearum". eng. In: *mBio* 8.5, e00895–17. ISSN: 2150-7511. DOI: 10.1128/mBio.00895-17.

Kiba, Akinori et al. (2007). "Pathogenesis-related protein 1 homologue is an antifungal protein in *Wasabia japonica* leaves and confers resistance to *Botrytis cinerea* in transgenic tobacco". In: *Plant Biotechnology* 24.2, pp. 247–253. DOI: 10.5511/plantbiotechnology.24.247.

Kim, Daehwan et al. (2016). "Centrifuge: rapid and sensitive classification of metagenomic sequences". eng. In: *Genome Research* 26.12, pp. 1721–1729. ISSN: 1549-5469. DOI: 10.1101/gr.210641.116.

Kim, Jin Yeong et al. (2014a). "Proteomics of rice and Cochliobolus miyabeanus fungal interaction: insight into proteins at intracellular and extracellular spaces". eng. In: *Proteomics* 14.20, pp. 2307–2318. ISSN: 1615-9861. DOI: 10.1002/pmic.201400066.

Kim, Moon Young et al. (2010). "Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 107.51, pp. 22032–22037. ISSN: 1091-6490. DOI: 10.1073/pnas.1009526107.

Kim, Myung-Shin et al. (2021). "Comparative analysis of de novo genomes reveals dynamic intra-species divergence of NLRs in pepper". In: *BMC Plant Biology* 21.1, p. 247. ISSN: 1471-2229. DOI: 10.1186/s12870-021-03057-8.

Kim, Sang Gon et al. (2013). "In-depth insight into in vivo apoplastic secretome of rice-Magnaporthe oryzae interaction". In: *Journal of Proteomics* 78, pp. 58–71. ISSN: 1874-3919. DOI: 10.1016/j.jprot.2012.10.029.

Kim, Seungill et al. (2014b). "Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species". en. In: *Nature Genetics* 46.3, pp. 270–278. ISSN: 1546-1718. DOI: 10.1038/ng.2877.

Klessig, Daniel F., Hyong Woo Choi, and D'Maris Amick Dempsey (2018). "Systemic acquired resistance and salicylic acid: Past, present, and future". eng. In: *Molecular*

plant-microbe interactions: MPMI 31.9, pp. 871–888. ISSN: 0894-0282. DOI: 10.1094/MPMI-03-18-0067-CR.

Kluyver, Thomas et al. (2016). "Jupyter Notebooks – a publishing format for reproducible computational workflows". en. In: *Jupyter*. Ed. by Fernando Loizides and Birgit Scmidt. IOS Press, pp. 87–90. DOI: 10.3233/978-1-61499-649-1-87.

Ko, Junsu and Juyong Lee (2021). "Can AlphaFold2 predict protein-peptide complex structures accurately?" en. In: Type: article, p. 2021.07.27.453972. DOI: https://doi.org/10.1101/2021.07.27.453972.

Konat, G., H. Offner, and J. Mellah (1984). "Improved sensitivity for detection and quantitation of glycoproteins on polyacrylamide gels". en. In: *Experientia* 40.3, pp. 303–304. ISSN: 1420-9071. DOI: 10.1007/BF01947594.

Kong, Qing et al. (2016). "Two redundant receptor-like cytoplasmic kinases function downstream of pattern recognition receptors to regulate activation of SA biosynthesis". In: *Plant Physiology* 171.2, pp. 1344–1354. ISSN: 0032-0889. DOI: 10.1104/pp.15.01954.

Kourelis, Jiorgos et al. (2019). "A homology-guided, genome-based proteome for improved proteomics in the alloploid Nicotiana benthamiana". In: *BMC Genomics* 20.1, p. 722. ISSN: 1471-2164. DOI: 10.1186/s12864-019-6058-6.

Kourelis, Jiorgos et al. (2020). "Evolution of a guarded decoy protease and its receptor in solanaceous plants". en. In: *Nature Communications* 11.1, p. 4393. ISSN: 2041-1723. DOI: 10.1038/s41467-020-18069-5.

Kryazhimskiy, Sergey and Joshua B. Plotkin (2008). "The Population Genetics of dN/dS". en. In: *PLOS Genetics* 4.12, e1000304. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1000304.

Kupriyanova, E., Polina Mamoshina, and Tatiana Ezhova (2015). "Evolutionary divergence of Arabidopsis thaliana classical peroxidases". In: *Biochemistry (Moscow)* 80. DOI: 10.1134/S0006297915100181.

Käll, Lukas et al. (2008). "Assigning significance to peptides identified by tandem mass spectrometry using decoy databases". In: *Journal of Proteome Research* 7.1, pp. 29–34. ISSN: 1535-3893. DOI: 10.1021/pr700600n.

Lam, Hon-Ming et al. (2010). "Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection". en. In: *Nature Genetics* 42.12, pp. 1053–1059. ISSN: 1546-1718. DOI: 10.1038/ng.715.

Lamb, Chris and Richard A. Dixon (1997). "The oxidative burst in plant disease resistance". In: *Annual Review of Plant Physiology and Plant Molecular Biology* 48.1, pp. 251–275. DOI: 10.1146/annurev.arplant.48.1.251.

Lane, B G (1994). "Oxalate, germin, and the extracellular matrix of higher plants". eng. In: *FASEB journal* 8.3, pp. 294–301. ISSN: 1530-6860. DOI: 10.1096/fasebj.8.3.8143935.

Laskowski, Michael et al. (1987). "Ovomucoid third domains from 100 avian species: isolation, sequences, and hypervariability of enzyme-inhibitor contact residues". In: *Biochemistry* 26.1, pp. 202–221. ISSN: 0006-2960. DOI: 10.1021/bi00375a028.

Laugé, R. et al. (1998). "Successful search for a resistance gene in tomato targeted against a virulence factor of a fungal pathogen". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 95.15, pp. 9014–9018. ISSN: 0027-8424. DOI: 10.1073/pnas.95.15.9014.

Laugé, R. et al. (2000). "Specific HR-associated recognition of secreted proteins from Cladosporium fulvum occurs in both host and non-host plants". eng. In: *The Plant Journal: For Cell and Molecular Biology* 23.6, pp. 735–745. ISSN: 0960-7412. DOI: 10.1046/j.1365-313x.2000.00843.x.

Laugé, Richard et al. (1997). "The in planta-produced Extracellular Proteins ECP1 and ECP2 of Cladosporium fulvum are virulence factors". In: *Molecular Plant-Microbe Interactions®* 10.6, pp. 725–734. ISSN: 0894-0282. DOI: 10.1094/MPMI.1997.10.6.725.

Lawton, Kay (1995). "Systemic acquired resistance in *Arabidopsis* requires salicylic acid but not ethylene". In: *Molecular Plant-Microbe Interactions* 8.6, p. 863. ISSN: 0894-0282. DOI: 10.1094/MPMI-8-0863.

Lazar, Cosmin et al. (2016). "Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies". In: *Journal of Proteome Research* 15.4, pp. 1116–1125. ISSN: 1535-3893. DOI: 10.1021/acs.jproteome.5b00981.

Lazic, Stanley E. (2010). "The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis?" In: *BMC Neuroscience* 11.1, p. 5. ISSN: 1471-2202. DOI: 10.1186/1471-2202-11-5.

Lee, Myoung-Hoon et al. (2019). "Lignin-based barrier restricts pathogens to the infection site and confers resistance in plants". In: *The EMBO Journal* 38.23, e101948. ISSN: 0261-4189. DOI: 10.15252/embj.2019101948.

Legrand, Michel et al. (1987). "Biological function of pathogenesis-related proteins: Four tobacco pathogenesis-related proteins are chitinases". In: *Proceedings of the National Academy of Sciences of the United States of America* 84.19, pp. 6750–6754. ISSN: 0027-8424. DOI: 10.1073/pnas.84.19.6750.

Leinonen, Rasko, Hideaki Sugawara, and Martin Shumway (2011). "The Sequence Read Archive". In: *Nucleic Acids Research* 39.Database issue, pp. D19–D21. ISSN: 0305-1048. DOI: 10.1093/nar/gkq1019.

Letunic, Ivica and Peer Bork (2021). "Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation". In: *Nucleic Acids Research* 49.W1, W293–W296. ISSN: 0305-1048. DOI: 10.1093/nar/gkab301.

Leubner-Metzger, G. et al. (1995). "Class I beta-1,3-glucanases in the endosperm of tobacco during germination". In: *Plant Physiology* 109.3, pp. 751–759. ISSN: 0032-0889. DOI: 10.1104/pp.109.3.751.

Li, Heng (2011). "A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data". In: *Bioinformatics* 27.21, pp. 2987–2993. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btr509.

— (2013). "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM". In: *arXiv:1303.3997 [q-bio]*. arXiv: 1303.3997. DOI: https://doi.org/10.48550/arXiv.1303.3997.

Li, Heng and Richard Durbin (2009). "Fast and accurate short read alignment with Burrows–Wheeler transform". In: *Bioinformatics* 25.14, pp. 1754–1760. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btp324.

Li, Heng et al. (2009). "The Sequence Alignment/Map format and SAMtools". In: *Bioinformatics* 25.16, pp. 2078–2079. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btp352.

Li, Qiang et al. (2020). "CsPrx25, a class III peroxidase in Citrus sinensis, confers resistance to citrus bacterial canker through the maintenance of ROS homeostasis and cell wall lignification". en. In: *Horticulture Research* 7.1, pp. 1–11. ISSN: 2052-7276. DOI: 10.1038/s41438-020-00415-9.

Li, Xiaodong et al. (2002). "Human receptors for sweet and umami taste". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 99.7, pp. 4692–4696. ISSN: 0027-8424. DOI: 10.1073/pnas.072090199.

Li, Ying-hui et al. (2014). "De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits". eng. In: *Nature Biotechnology* 32.10, pp. 1045–1052. ISSN: 1546-1696. DOI: 10.1038/nbt.2979.

Li, Youshan et al. (2016). "Loss of second and sixth conserved cysteine residues from trypsin inhibitor-like cysteine-rich domain-type protease inhibitors in Bombyx mori may induce activity against microbial proteases". eng. In: *Peptides* 86, pp. 13–23. ISSN: 1873-5169. DOI: 10.1016/j.peptides.2016.09.013.

Liigand, Piia, Karl Kaupmees, and Anneli Kruve (2019). "Influence of the amino acid composition on the ionization efficiencies of small peptides". eng. In: *Journal of mass spectrometry: JMS* 54.6, pp. 481–487. ISSN: 1096-9888. DOI: 10.1002/jms.4348.

Lin, Tao et al. (2014). "Genomic analyses provide insights into the history of tomato breeding". en. In: *Nature Genetics* 46.11, pp. 1220–1226. ISSN: 1546-1718. DOI: 10.1038/ng.3117.

Liu, Diqiu et al. (2012). "Molecular cloning of a thaumatin-like protein gene from Pyrus pyrifolia and overexpression of this gene in tobacco increased resistance to pathogenic fungi". en. In: *Plant Cell, Tissue and Organ Culture (PCTOC)* 111.1, pp. 29–39. ISSN: 1573-5044. DOI: 10.1007/s11240-012-0167-0.

Liu, Jun-Jun, Rona Sturrock, and Abul K. M. Ekramoddoullah (2010). "The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function". en. In: *Plant Cell Reports* 29.5, pp. 419–436. ISSN: 1432-203X. DOI: 10.1007/s00299-010-0826-8.

Liu, Qiong et al. (2018). "Assembly and annotation of a draft genome sequence for Glycine latifolia, a perennial wild relative of soybean". eng. In: *The Plant Journal: For Cell and Molecular Biology* 95.1, pp. 71–85. ISSN: 1365-313X. DOI: 10.1111/tpj.13931.

Liu, Yongsheng, Matthew P. Patricelli, and Benjamin F. Cravatt (1999). "Activity-based protein profiling: the serine hydrolases". In: *Proceedings of the National Academy of Sciences* 96.26, pp. 14694–14699. DOI: 10.1073/pnas.96.26.14694.

Liu, Yucheng et al. (2020). "Pan-Genome of Wild and Cultivated Soybeans". en. In: *Cell* 182.1, 162–176.e13. ISSN: 0092-8674. DOI: 10.1016/j.cell.2020.05.023.

Llorente, Francisco et al. (2002). "A novel cold-inducible gene from Arabidopsis, RCI3, encodes a peroxidase that constitutes a component for stress tolerance". eng. In: *The Plant Journal: For Cell and Molecular Biology* 32.1, pp. 13–24. ISSN: 0960-7412. DOI: 10.1046/j.1365-313x.2002.01398.x.

Lozano-Torres, Jose L. et al. (2012). "Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode". en. In: *Proceedings of the National Academy of Sciences* 109.25, pp. 10119–10124. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1202867109.

Lu, Stephen M. et al. (2001). "Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases". en. In: *Proceedings of the National Academy of Sciences* 98.4, pp. 1410–1415. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.98.4.1410.

Lu, W. et al. (1997). "Binding of amino acid side-chains to S1 cavities of serine proteinases". eng. In: *Journal of Molecular Biology* 266.2, pp. 441–461. ISSN: 0022-2836. DOI: 10.1006/jmbi.1996.0781.

Lucas, John A., Nichola J. Hawkins, and Bart A. Fraaije (2015). "Chapter two - The evolution of fungicide resistance". en. In: *Advances in Applied Microbiology*. Ed. by SIMA Sariaslani and GEOFFREY MICHAEL Gadd. Vol. 90. Academic Press, pp. 29–92. DOI: 10.1016/bs.aambs.2014.09.001.

Lyr, H. et al., eds. (1999). *Modern fungicides and antifungal compounds II*. English. 1st edition. Andover: Intercept. ISBN: 978-1-898298-60-1.

Ma, Li-Jun et al. (2010). "Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium". eng. In: *Nature* 464.7287, pp. 367–373. ISSN: 1476-4687. DOI: 10.1038/nature08850.

Ma, Lisong et al. (2015). "The AVR2-SIX5 gene pair is required to activate I-2-mediated immunity in tomato". eng. In: *The New Phytologist* 208.2, pp. 507–518. ISSN: 1469-8137. DOI: 10.1111/nph.13455.

Ma, Zhenchuan et al. (2017). "A paralogous decoy protects Phytophthora sojae apoplastic effector PsXEG1 from a host inhibitor". en. In: *Science* 355.6326, pp. 710–714. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aai7919.

Manohar, Murli et al. (2015). "Identification of multiple salicylic acid-binding proteins using two high throughput screens". In: *Frontiers in Plant Science* 5. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2014.00777.

Mansfield, John et al. (2012). "Top 10 plant pathogenic bacteria in molecular plant pathology". en. In: *Molecular Plant Pathology* 13.6, pp. 614–629. ISSN: 1364-3703. DOI: 10.1111/j.1364-3703.2012.00804.x.

Mariani, Valerio et al. (2013). "lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests". In: *Bioinformatics* 29.21, pp. 2722–2728. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btt473.

Martin, Marcel et al. (2016). "WhatsHap: fast and accurate read-based phasing". en. In: *bioRxiv*, p. 085050. DOI: 10.1101/085050.

Martínez-Cruz, Jesús et al. (2021). "Effectors with chitinase activity (EWCAs), a family of conserved, secreted fungal chitinases that suppress chitin-triggered immunity". In: *The Plant Cell* 33.4, pp. 1319–1340. ISSN: 1040-4651. DOI: 10.1093/plcell/koab011.

Martínez-González, A. P. et al. (2018). "What proteomic analysis of the apoplast tells us about plant–pathogen interactions". en. In: *Plant Pathology* 0.0. ISSN: 1365-3059. DOI: 10.1111/ppa.12893.

Matsumoto, Ayumi et al. (2022). "A versatile Tn7 transposon-based bioluminescence tagging tool for quantitative and spatial detection of bacteria in plants". en. In: *Plant Communications* 3.1, p. 100227. ISSN: 2590-3462. DOI: 10.1016/j.xplc.2021.100227.

McDowell, J M et al. (1998). "Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsis." In: *The Plant Cell* 10.11, pp. 1861–1874. ISSN: 1040-4651. DOI: 10.1105/tpc.10.11.1861.

Melin, Petter, Johan Schnürer, and E. Gerhart H. Wagner (2003). "Characterization of phiA, a gene essential for phialide development in Aspergillus nidulans". eng. In: *Fungal genetics and biology: FG & B* 40.3, pp. 234–241. ISSN: 1087-1845. DOI: 10.1016/s1087-1845(03)00108-7.

Menu-Bouaouiche, Laurence et al. (2003). "A molecular basis for the endo-1,3-glucanase activity of the thaumatin-like proteins from edible fruits". en. In: *Biochimie*. Free and conjugated sugars. Biochemistry and molecular biology. Dedicated to Andre Verbert and Jean Agneray 85.1, pp. 123–131. ISSN: 0300-9084. DOI: 10.1016/S0300-9084(03)00058-0.

Mesarich, Carl H. et al. (2018). "Specific hypersensitive response-associated recognition of new apoplastic effectors from Cladosporium fulvum in wild tomato". eng. In: *Molecular plant-microbe interactions: MPMI* 31.1, pp. 145–162. ISSN: 0894-0282. DOI: 10.1094/MPMI-05-17-0114-FI.

Meyer, Michael et al. (2016). "The subtilisin-like protease SBT3 contributes to insect resistance in tomato". In: *Journal of Experimental Botany* 67.14, pp. 4325–4338. ISSN: 0022-0957. DOI: 10.1093/jxb/erw220.

Min, Kyeongsik et al. (2004). "Crystal structure of osmotin, a plant antifungal protein". en. In: *Proteins: Structure, Function, and Bioinformatics* 54.1, pp. 170–173. ISSN: 1097-0134. DOI: 10.1002/prot.10571.

Mirdita, Milot et al. (2022). "ColabFold: making protein folding accessible to all". en. In: *Nature Methods* 19.6, pp. 679–682. ISSN: 1548-7105. DOI: 10.1038/s41592-022-01488-1.

Misas-Villamil, Johana C. et al. (2013). "Pseudomonas syringae pv. syringae uses proteasome inhibitor Syringolin A to colonize from wound infection sites". en. In: *PLOS Pathogens* 9.3, e1003281. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1003281.

Misra, Rajesh Chandra et al. (2016). "A thaumatin-like protein of Ocimum basilicum confers tolerance to fungal pathogen and abiotic stress in transgenic Arabidopsis". en. In: *Scientific Reports* 6.1, p. 25340. ISSN: 2045-2322. DOI: 10.1038/srep25340.

Monteiro, Filipa et al. (2018). "Genetic and genomic tools to asssist sugar beet improvement: the value of the crop wild relatives". In: *Frontiers in Plant Science* 9. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2018.00074.

Morimoto, Kyoko et al. (2022). "Broad-range metalloprotease profiling in plants uncovers immunity provided by defence-related metalloenzyme". en. In: *New Phytologist* n/a.n/a. ISSN: 1469-8137. DOI: 10.1111/nph.18200.

Mueller, André N. et al. (2013). "Compatibility in the Ustilago maydis–maize interaction requires inhibition of host cysteine proteases by the fungal effector Pit2". In: *PLoS Pathogens* 9.2. ISSN: 1553-7366. DOI: 10.1371/journal.ppat.1003177.

Mölder, Felix et al. (2021). "Sustainable data analysis with Snakemake". en. In: Type: article. DOI: https://doi.org/10.12688/f1000research.29032.2.

Müller, Nathalie et al. (2018). "Investigations on VELVET regulatory mutants confirm the role of host tissue acidification and secretion of proteins in the pathogenesis of Botrytis cinerea". en. In: *New Phytologist* 219.3, pp. 1062–1074. ISSN: 1469-8137. DOI: 10.1111/nph.15221.

Müller, Torsten and Dominic Winter (2017). "Systematic evaluation of protein reduction and alkylation reveals massive unspecific side effects by iodine-containing reagents". In: *Molecular & Cellular Proteomics : MCP* 16.7, pp. 1173–1187. ISSN: 1535-9476. DOI: 10.1074/mcp.M116.064048.

Naumann, Todd A., Donald T. Wicklow, and Neil P. J. Price (2011). "Identification of a chitinase-modifying protein from Fusarium verticillioides. Truncation of a host resistance protein by a fungalysin metalloprotease". en. In: *Journal of Biological Chemistry* 286.41, pp. 35358–35366. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M111.279646.

Navarre, D. A. and D. Mayo (2004). "Differential characteristics of salicylic acid-mediated signaling in potato". en. In: *Physiological and Molecular Plant Pathology* 64.4, pp. 179–188. ISSN: 0885-5765. DOI: 10.1016/j.pmpp.2004.09.001.

Nelson, Rebecca et al. (2018). "Navigating complexity to breed disease-resistant crops". en. In: *Nature Reviews Genetics* 19.1, pp. 21–33. ISSN: 1471-0064. DOI: 10.1038/nrg.2017.82.

Ngou, Bruno Pok Man, Pingtao Ding, and Jonathan D G Jones (2022). "Thirty years of resistance: Zig-zag through the plant immune system". In: *The Plant Cell*, koac041. ISSN: 1040-4651. DOI: 10.1093/plcell/koac041.

Ngou, Bruno Pok Man et al. (2021). "Mutual potentiation of plant immunity by cell-surface and intracellular receptors". en. In: *Nature* 592.7852, pp. 110–115. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03315-7.

Nie, Jiajun et al. (2022). "Two NIS1-like proteins from apple canker pathogen (Valsa mali) play distinct roles in plant recognition and pathogen virulence". en. In: *Stress Biology* 2.1, p. 7. ISSN: 2731-0450. DOI: 10.1007/s44154-021-00031-0.

Nielsen, Rasmus and Ziheng Yang (1998). "Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene". en. In: *Genetics* 148.3, pp. 929–936. ISSN: 0016-6731, 1943-2631. DOI: https://doi.org/10.1093/genetics/148.3.929.

Nishimura, Marc T. and Jeffery L. Dangl (2010). "Arabidopsis and the plant immune system". eng. In: *The Plant Journal: For Cell and Molecular Biology* 61.6, pp. 1053–1066. ISSN: 1365-313X. DOI: 10.1111/j.1365-313X.2010.04131.x.

Nobori, Tatsuya et al. (2020). "Multidimensional gene regulatory landscape of a bacterial pathogen in plants". eng. In: *Nature Plants* 6.7, pp. 883–896. ISSN: 2055-0278. DOI: 10.1038/s41477-020-0690-7.

O'Leary, Brendan M. et al. (2014). "The infiltration-centrifugation technique for extraction of apoplastic fluid from plant leaves using Phaseolus vulgaris as an example". In: *Journal of Visualized Experiments : JoVE* 12.94. ISSN: 1940-087X. DOI: 10.3791/52113.

Ondov, Brian D. et al. (2016). "Mash: fast genome and metagenome distance estimation using MinHash". In: *Genome Biology* 17.1, p. 132. ISSN: 1474-760X. DOI: 10.1186/s13059-016-0997-x.

Ong, Shao-En et al. (2002). "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics". eng. In: *Molecular & cellular proteomics: MCP* 1.5, pp. 376–386. ISSN: 1535-9476. DOI: 10.1074/mcp.m200025-mcp200.

Ottmann, Christian et al. (2009). "Structural basis for Ca2+-independence and activation by homodimerization of tomato subtilase 3". In: *Proceedings of the National Academy of Sciences* 106.40, pp. 17223–17228. DOI: 10.1073/pnas.0907587106.

Paniagua, Candelas, Louisa Perry, and Yoselin Benitez-Alfonso (2021). "A phylogenetic and transcriptomic study of the -1,3-glucanase family in tomato identifies candidate targets for fruit improvement". en. In: Type: article, p. 2021.09.29.462359. DOI: https://doi.org/10.1101/2021.09.29.462359.

Panstruga, Ralph and Matthew J. Moscou (2020). "What is the molecular basis of nonhost resistance?" In: *Molecular Plant-Microbe Interactions®* 33.11, pp. 1253–1264. ISSN: 0894-0282. DOI: 10.1094/MPMI-06-20-0161-CR.

Parent, Jean-Guy, Richard Hogue, and Alain Asselin (1985). "Glycoproteins, enzymatic activities, and b proteins in intercellular fluid extracts from hypersensitive Nicotiana species infected with tobacco mosaic virus". In: *Canadian Journal of Botany* 63.5, pp. 928–931. ISSN: 0008-4026. DOI: 10.1139/b85-123.

Parker, Jennifer et al. (2013). "Quantitative proteomics of tomato defense against Pseudomonas syringae infection". eng. In: *Proteomics* 13.12-13, pp. 1934–1946. ISSN: 1615-9861. DOI: 10.1002/pmic.201200402.

Passardi, Filippo et al. (2004). "The class III peroxidase multigenic family in rice and its evolution in land plants". eng. In: *Phytochemistry* 65.13, pp. 1879–1893. ISSN: 0031-9422. DOI: 10.1016/j.phytochem.2004.06.023.

Passardi, Filippo et al. (2006). "Two cell wall associated peroxidases from Arabidopsis influence root elongation". eng. In: *Planta* 223.5, pp. 965–974. ISSN: 0032-0935. DOI: 10.1007/s00425-005-0153-4.

Patel, Jignasha et al. (2007). "Elastin-like polypeptide fusions enhance the accumulation of recombinant proteins in tobacco leaves". en. In: *Transgenic Research* 16.2, pp. 239–249. ISSN: 1573-9368. DOI: 10.1007/s11248-006-9026-2.

Patricelli, M. P. et al. (2001). "Direct visualization of serine hydrolase activities in complex proteomes using fluorescent active site-directed probes". eng. In: *Proteomics* 1.9, pp. 1067–1071. ISSN: 1615-9853. DOI: 10.1002/1615-9861(200109)1:9<1067::AID-PROT1067>3.0.CO;2-4.

Paulus, Judith K. et al. (2020). "Extracellular proteolytic cascade in tomato activates immune protease Rcr3". en. In: *Proceedings of the National Academy of Sciences* 117.29, pp. 17409–17417. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1921101117.

Pedregosa, Fabian et al. (2011). "Scikit-learn: Machine learning in Python". In: *Journal of Machine Learning Research* 12.85, pp. 2825–2830. ISSN: 1533-7928. DOI: https://doi.org/10.48550/arXiv.1201.0490.

Pei, Yakun et al. (2019). "GhABP19, a novel germin-like protein from Gossypium hirsutum, plays an important role in the regulation of resistance to Verticillium and Fusarium wilt pathogens". In: *Frontiers in Plant Science* 10. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2019.00583.

Peralta, I. E., D. M. Spooner, and S. Knapp (2008). "Taxonomy of wild tomatoes and their relatives (Solanum sect. Lycopersicoides, sect. Juglandifolia, sect. Lycopersicon; Solanaceae)." English. In: *Systematic Botany Monographs* 84. ISSN: 0737-8211.

Perrot, Thomas, Markus Pauly, and Vicente Ramírez (2022). "Emerging roles of -glucanases in plant development and adaptative responses". en. In: *Plants* 11.9, p. 1119. ISSN: 2223-7747. DOI: 10.3390/plants11091119.

Peterman, Scott M. and Joseph J. Mulholland (2006). "A novel approach for identification and characterization of glycoproteins using a hybrid linear Ion Trap/FT-ICR mass spectrometer". en. In: *Journal of the American Society for Mass Spectrometry* 17.2, pp. 168–179. ISSN: 1044-0305. DOI: 10.1016/j.jasms.2005.10.008.

Pieterse, Corné M. J. et al. (2009). "Networking by small-molecule hormones in plant immunity". en. In: *Nature Chemical Biology* 5.5, pp. 308–316. ISSN: 1552-4469. DOI: 10.1038/nchembio.164.

Planas-Marquès, Marc et al. (2018). "Protease activities triggered by Ralstonia solanacearum infection in susceptible and tolerant tomato Llines". en. In: *Molecular & Cellular Pro-*

teomics 17.6, pp. 1112–1125. ISSN: 1535-9476, 1535-9484. DOI: 10.1074/mcp.RA117.000052.

Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin (2010). "FastTree 2 – Approximately maximum-likelihood trees for large alignments". en. In: *PLOS ONE* 5.3, e9490. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0009490.

Punja, Zamir K. and Ye-Yan Zhang (1993). "Plant chitinases and their roles in resistance to fungal diseases". In: *Journal of Nematology* 25.4, pp. 526–540. ISSN: 0022-300X.

Pusztahelyi, Tünde (2018). "Chitin and chitin-related compounds in plant–fungal interactions". In: *Mycology* 9.3, pp. 189–201. ISSN: 2150-1203. DOI: 10.1080/21501203.2018.1473299.

Qiao, Qin et al. (2021). "Evolutionary history and pan-genome dynamics of strawberry (Fragaria spp.)" In: *Proceedings of the National Academy of Sciences* 118.45, e2105431118. DOI: 10.1073/pnas.2105431118.

Ramírez, Vicente et al. (2013). "An extracellular subtilase switch for immune priming in Arabidopsis". eng. In: *PLoS pathogens* 9.6, e1003445. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1003445.

Ranallo-Benavidez, T. Rhyker, Kamil S. Jaron, and Michael C. Schatz (2020). "GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes". en. In: *Nature Communications* 11.1, p. 1432. ISSN: 2041-1723. DOI: 10.1038/s41467-020-14998-3.

Ravi Kumar, Majeti N. V (2000). "A review of chitin and chitosan applications". en. In: *Reactive and Functional Polymers* 46.1, pp. 1–27. ISSN: 1381-5148. DOI: 10.1016/S1381-5148(00)00038-9.

Rawlings, Neil D. et al. (2018). "The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database". en. In: *Nucleic Acids Research* 46.D1, pp. D624–D632. ISSN: 0305-1048. DOI: 10.1093/nar/gkx1134.

Razali, Rozaimi et al. (2018). "The genome sequence of the wild tomato Solanum pimpinellifolium provides insights into salinity tolerance". In: *Frontiers in Plant Science* 9. ISSN: 1664-462X. DOI: 10.3389/fpls.2018.01402.

Read, R. J. et al. (1983). "Structure of the complex of Streptomyces griseus protease B and the third domain of the turkey ovomucoid inhibitor at 1.8-A resolution". eng. In: *Biochemistry* 22.19, pp. 4420–4433. ISSN: 0006-2960. DOI: https://doi.org/10.1021/bi00288a012.

Reichardt, Sven et al. (2018). "The tomato subtilase family includes several cell death-related proteinases with caspase specificity". en. In: *Scientific Reports* 8.1, p. 10531. ISSN: 2045-2322. DOI: 10.1038/s41598-018-28769-0.

Rep, Martijn et al. (2004). "A small, cysteine-rich protein secreted by Fusarium oxysporum during colonization of xylem vessels is required for I-3-mediated resistance in tomato". eng. In: *Molecular Microbiology* 53.5, pp. 1373–1383. ISSN: 0950-382X. DOI: 10.1111/j.1365-2958.2004.04177.x.

Richards, Andrew (2015). *University of Oxford advanced research computing*. DOI: 10.5281/zenodo.22558.

Roberts, Richard J. (2018). "The Nobel Laureates' Campaign Supporting GMOs". en. In: *Journal of Innovation & Knowledge*. Poverty, Hunger and Sustainability 3.2, pp. 61–65. ISSN: 2444-569X. DOI: 10.1016/j.jik.2017.12.006.

Roberts, Walden K. and Claude P.YR 1990 Selitrennikoff (1990). "Zeamatin, an antifungal protein from maize with membrane-permeabilizing activity". In: *Microbiology* 136.9, pp. 1771–1778. ISSN: 1465-2080, DOI: 10.1099/00221287-136-9-1771.

Romera, Francisco J. et al. (2019). "Induced systemic resistance (ISR) and Fe deficiency responses in dicot plants". In: *Frontiers in Plant Science* 10. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2019.00287.

Romero, Irene et al. (2008). "Functionality of a class I beta-1,3-glucanase from skin of table grapes berries". en. In: *Plant Science* 174.6, pp. 641–648. ISSN: 0168-9452. DOI: 10.1016/j.plantsci.2008.03.019.

Rooney, Henrietta C. et al. (2005). "Cladosporium avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance". en. In: *Science* 308.5729, pp. 1783–1786. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1111404.

Rose, Jocelyn K. C. et al. (2002). "Molecular cloning and characterization of glucanase inhibitor proteins: coevolution of a counterdefense mechanism by plant pathogens". eng. In: *The Plant Cell* 14.6, pp. 1329–1345. ISSN: 1040-4651. DOI: 10.1105/tpc.002253.

Rosengren, K. Johan et al. (2001). "Solution structure of BSTI: a new trypsin inhibitor from skin secretions of bombina bombina". In: *Biochemistry* 40.15, pp. 4601–4609. ISSN: 0006-2960. DOI: 10.1021/bi002623v.

Ross, A. Frank (1961). "Systemic acquired resistance induced by localized virus infections in plants". en. In: *Virology* 14.3, pp. 340–358. ISSN: 0042-6822. DOI: 10.1016/0042-6822(61)90319-1.

Salanoubat, M. et al. (2002). "Genome sequence of the plant pathogen Ralstonia solanacearum". eng. In: *Nature* 415.6871, pp. 497–502. ISSN: 0028-0836. DOI: 10.1038/415497a.

Sarowar, Sujon et al. (2005). "Overexpression of a pepper basic pathogenesis-related protein 1 gene in tobacco plants enhances resistance to heavy metal and pathogen stresses". en. In: *Plant Cell Reports* 24.4, pp. 216–224. ISSN: 1432-203X. DOI: 10.1007/s00299-005-0928-x.

Sasaki, Shinya et al. (2004). "Lignin dehydrogenative polymerization mechanism: a poplar cell wall peroxidase directly oxidizes polymer lignin and produces in vitro dehydrogenative polymer rich in beta-O-4 linkage". eng. In: *FEBS letters* 562.1-3, pp. 197–201. ISSN: 0014-5793. DOI: 10.1016/S0014-5793(04)00224-8.

Savary, Serge et al. (2019). "The global burden of pathogens and pests on major food crops". eng. In: *Nature Ecology & Evolution* 3.3, pp. 430–439. ISSN: 2397-334X. DOI: 10.1038/s41559-018-0793-y.

Schlumbaum, Angela et al. (1986). "Plant chitinases are potent inhibitors of fungal growth". en. In: *Nature* 324.6095, pp. 365–367. ISSN: 1476-4687. DOI: 10.1038/324365a0.

Schmidt, Sarah M. et al. (2013). "MITEs in the promoters of effector genes allow prediction of novel virulence genes in Fusarium oxysporum". eng. In: *BMC genomics* 14, p. 119. ISSN: 1471-2164. DOI: 10.1186/1471-2164-14-119.

Schmutz, Jeremy et al. (2010). "Genome sequence of the palaeopolyploid soybean". eng. In: *Nature* 463.7278, pp. 178–183. ISSN: 1476-4687. DOI: 10.1038/nature08670.

Schneider, Caroline A., Wayne S. Rasband, and Kevin W. Eliceiri (2012). "NIH Image to ImageJ: 25 years of image analysis". eng. In: *Nature Methods* 9.7, pp. 671–675. ISSN: 1548-7105. DOI: https://doi.org/10.1038/nmeth.2089.

Schnippenkoetter, Wendelin et al. (2021). "Comparison of non-subjective relative fungal biomass measurements to quantify the Leptosphaeria maculans—Brassica napus interaction". In: *Plant Methods* 17.1, p. 122. ISSN: 1746-4811. DOI: 10.1186/s13007-021-00822-6.

Schouten, Henk J. et al. (2019). "Breeding has increased the diversity of cultivated tomato in the Netherlands". In: *Frontiers in Plant Science* 10. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2019.01606.

Sedivy, John M. (2011). "Phosphatidylenthanolamine binding protein aka Raf kinase inhibitor protein: a brief history of its discovery and the remarkable diversity of biological functions". In: *Forum on immunopathological diseases and therapeutics* 2.1, pp. 1–12. ISSN: 2151-8017. DOI: 10.1615/ForumImmunDisTher.v2.i1.20.

Serrano, Irene et al. (2016). "A non canonical subtilase attenuates the transcriptional activation of defence responses in Arabidopsis thaliana". In: *eLife* 5. Ed. by Gary Stacey, e19755. ISSN: 2050-084X. DOI: 10.7554/eLife.19755.

Serrano, Mario et al. (2014). "The cuticle and plant defense to pathogens". In: *Frontiers in Plant Science* 5, p. 274. ISSN: 1664-462X. DOI: 10.3389/fpls.2014.00274.

Seyfferth, Carolin and Kenichi Tsuda (2014). "Salicylic acid signal transduction: the initiation of biosynthesis, perception and transcriptional reprogramming". In: *Frontiers in Plant Science* 5. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2014.00697.

Shabab, Mohammed et al. (2008). "Fungal effector protein AVR2 targets diversifying defense-related cys P proteases of tomato". In: *The Plant Cell* 20.4, pp. 1169–1183. ISSN: 1040-4651. DOI: 10.1105/tpc.107.056325.

Shah, Jyoti, Frank Tsui, and Daniel F. Klessig (1997). "Characterization of a salicylic acid-insensitive mutant (sai1) of Arabidopsis thaliana, identified in a selective screen utilizing the SA-inducible expression of the tms2 gene". In: *Molecular Plant-Microbe Interactions®* 10.1, pp. 69–78. ISSN: 0894-0282. DOI: 10.1094/MPMI.1997.10.1.69.

Shao, Dandan et al. (2021). "Effectors of plant necrotrophic fungi". In: *Frontiers in Plant Science* 12. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2021.687713.

Shen, Wei et al. (2016). "SeqKit: a cross-platform and ultrafast toolkit for fastas/q file manipulation". en. In: *PLOS ONE* 11.10, e0163962. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0163962.

Sherman, Rachel M. and Steven L. Salzberg (2020). "Pan-genomics in the human genome era". en. In: *Nature Reviews Genetics* 21.4, pp. 243–254. ISSN: 1471-0064. DOI: 10.1038/s41576-020-0210-7.

Shigeto, Jun and Yuji Tsutsumi (2016). "Diverse functions and reactions of class III peroxidases". en. In: *New Phytologist* 209.4, pp. 1395–1402. ISSN: 1469-8137. DOI: 10.1111/nph.13738.

Shigeto, Jun et al. (2015). "Simultaneously disrupting AtPrx2, AtPrx25 and AtPrx71 alters lignin content and structure in Arabidopsis stem". en. In: *Journal of Integrative Plant Biology* 57.4, pp. 349–356. ISSN: 1744-7909. DOI: 10.1111/jipb.12334.

Shin, Sang Hyun et al. (2014). "An acidic pathogenesis-related 1gene of Oryza grandiglumis is involved in disease resistance response against bacterial infection". In: *The Plant Pathology Journal* 30.2, pp. 208–214. ISSN: 1598-2254. DOI: 10.5423/PPJ.NT.11.2013.0112.

Shindo, Takayuki et al. (2016). "Screen of non-annotated small secreted proteins of Pseudomonas syringae reveals a virulence factor that inhibits tomato immune proteases". eng. In: *PLoS pathogens* 12.9, e1005874. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1005874.

Sierro, Nicolas et al. (2014). "The tobacco genome sequence and its comparison with those of tomato and potato". en. In: *Nature Communications* 5.1, p. 3833. ISSN: 2041-1723. DOI: 10.1038/ncomms4833.

Simbaqueba, Jaime et al. (2018). "Evidence for horizontal gene transfer and separation of effector recognition from effector function revealed by analysis of effector genes shared between cape gooseberry- and tomato-infecting formae speciales of Fusarium oxysporum". In: *Molecular Plant Pathology* 19.10, pp. 2302–2318. ISSN: 1464-6722. DOI: 10.1111/mpp.12700.

Simbaqueba, Jaime et al. (2021). "Putative novel effector genes revealed by the genomic analysis of the phytopathogenic fungus Fusarium oxysporum f. sp. physali (Foph) that infects cape gooseberry plants". In: *Frontiers in Microbiology* 11. ISSN: 1664-302X. DOI: https://doi.org/10.3389/fmicb.2020.593915.

Simão, Felipe A. et al. (2015). "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs". In: *Bioinformatics* 31.19, pp. 3210–3212. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btv351.

Sinha, Mau et al. (2014). "Current overview of allergens of plant pathogenesis-related protein families". en. In: *The Scientific World Journal* 2014, e543195. ISSN: 2356-6140. DOI: 10.1155/2014/543195.

Slater, Guy St C. and Ewan Birney (2005). "Automated generation of heuristics for biological sequence comparison". eng. In: *BMC bioinformatics* 6, p. 31. ISSN: 1471-2105. DOI: 10.1186/1471-2105-6-31.

Soldan, Riccardo et al. (2020). "From macro to micro: a combined bioluminescence-fluorescence approach to monitor bacterial localization". en. In: *Environmental Microbiology* n/a.n/a. ISSN: 1462-2920. DOI: https://doi.org/10.1111/1462-2920.15296.

Soltis, Nicole E. et al. (2019). "Interactions of tomato and Botrytis cinerea genetic diversity: Parsing the contributions of host differentiation, domestication, and pathogen variation". en. In: *The Plant Cell* 31.2, pp. 502–519. ISSN: 1040-4651, 1532-298X. DOI: 10.1105/tpc.18.00857.

Song, Jing et al. (2009). "Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3". en. In: *Proceedings of the National Academy of Sciences* 106.5, pp. 1654–1659. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0809201106.

Sonnhammer, E. L., G. von Heijne, and A. Krogh (1998). "A hidden Markov model for predicting transmembrane helices in protein sequences". eng. In: *Proceedings. International Conference on Intelligent Systems for Molecular Biology* 6, pp. 175–182. ISSN: 1553-0833.

Sperschneider, Jana et al. (2018). "ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning". en. In: *New Phytologist* 217.4, pp. 1764–1778. ISSN: 1469-8137. DOI: 10.1111/nph.14946.

Stael, Simon et al. (2015). "Plant innate immunity – sunny side up?" en. In: *Trends in Plant Science* 20.1, pp. 3–11. ISSN: 1360-1385. DOI: 10.1016/j.tplants.2014.10.002.

Stahl, Eli A. et al. (1999). "Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*". en. In: *Nature* 400.6745, pp. 667–671. ISSN: 1476-4687. DOI: 10.1038/23260.

Stam, Remco, Gustavo A. Silva-Arias, and Aurelien Tellier (2019a). "Subsets of NLR genes show differential signatures of adaptation during colonization of new habitats". en. In: *New Phytologist* 224.1, pp. 367–379. ISSN: 1469-8137. DOI: 10.1111/nph.16017.

Stam, Remco et al. (2019b). "The de novo reference genome and transcriptome assemblies of the wild tomato species Solanum chilense highlights birth and death of NLR genes between tomato species". en. In: *G3: Genes, Genomes, Genetics* 9.12, pp. 3933–3941. ISSN: 2160-1836. DOI: 10.1534/g3.119.400529.

Stamatakis, Alexandros (2014). "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies". en. In: *Bioinformatics* 30.9, pp. 1312–1313. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btu033.

Stintzi, A. et al. (1993). "Plant 'pathogenesis-related' proteins and their role in defense against pathogens". eng. In: *Biochimie* 75.8, pp. 687–706. ISSN: 0300-9084. DOI: 10.1016/0300-9084(93)90100-7.

Stubbs, M T et al. (1990). "The refined 2.4 A X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction." In: *The EMBO Journal* 9.6, pp. 1939–1947. ISSN: 0261-4189. DOI: 10.1002/j.1460-2075.1990.tb08321.x.

Student (1908). "The probable error of a mean". In: *Biometrika* 6.1, pp. 1–25. ISSN: 0006-3444. DOI: 10.2307/2331554.

Sueldo, Daniela et al. (2014). "Dynamic hydrolase activities precede hypersensitive tissue collapse in tomato seedlings". en. In: *New Phytologist* 203.3, pp. 913–925. ISSN: 1469-8137. DOI: 10.1111/nph.12870.

Suyama, Mikita, David Torrents, and Peer Bork (2006). "PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments". In: *Nucleic Acids Research* 34.Web Server issue, W609–W612. ISSN: 0305-1048. DOI: 10.1093/nar/gkl315.

Svanström, Åsa et al. (2013). "The lactic acid bacteria metabolite phenyllactic acid inhibits both radial growth and sporulation of filamentous fungi". In: *BMC Research Notes* 6.1, p. 464. ISSN: 1756-0500. DOI: 10.1186/1756-0500-6-464.

Sánchez-Vallet, Andrea et al. (2013). "Fungal effector Ecp6 outcompetes host immune receptor for chitin binding through intrachain LysM dimerization". In: *eLife* 2. Ed. by Jean Greenberg, e00790. ISSN: 2050-084X. DOI: 10.7554/eLife.00790.

Takashima, Tomoya et al. (2020). "Antifungal activities of LysM-domain multimers and their fusion chitinases". eng. In: *International Journal of Biological Macromolecules* 154, pp. 1295–1302. ISSN: 1879-0003. DOI: 10.1016/j.ijbiomac.2019.11.005.

Takatsuji, Hiroshi (2014). "Development of disease-resistant rice using regulatory components of induced disease resistance". In: *Frontiers in Plant Science* 5. ISSN: 1664-462X. DOI: https://doi.org/10.3389/fpls.2014.00630.

Takemoto, D. et al. (1997). "Identification of chitinase and osmotin-like protein as actin-binding proteins in suspension-cultured potato cells". eng. In: *Plant & Cell Physiology* 38.4, pp. 441–448. ISSN: 0032-0781. DOI: 10.1093/oxfordjournals.pcp.a029187.

Tang, Dié et al. (2022). "Genome evolution and diversity of wild and cultivated potatoes". en. In: *Nature* 606.7914. Number: 7914 Publisher: Nature Publishing Group, pp. 535–541. ISSN: 1476-4687. DOI: 10.1038/s41586-022-04822-x.

Tang, Muya et al. (2019). "Functional analysis and characterization of antimicrobial phosphatidylethanolamine-binding protein BmPEBP in the silkworm Bombyx mori". eng. In: *Insect Biochemistry and Molecular Biology* 110, pp. 1–9. ISSN: 1879-0240. DOI: 10.1016/j.ibmb.2019.03.011.

Taylor, Alexander and Yin-Long Qiu (2017). "Evolutionary history of subtilases in land plants and their involvement in symbiotic interactions". eng. In: *Molecular plant-microbe interactions: MPMI* 30.6, pp. 489–501. ISSN: 0894-0282. DOI: 10.1094/MPMI-10-16-0218-R.

Ten Have, Arjen et al. (2007). "Partial stem and leaf resistance against the fungal pathogen Botrytis cinerea in wild relatives of tomato". en. In: *European Journal of Plant Pathology* 117.2, pp. 153–166. ISSN: 1573-8469. DOI: 10.1007/s10658-006-9081-9.

Terwisscha van Scheltinga, A. C. et al. (1995). "Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: evidence for substrate assisted catalysis". eng. In: *Biochemistry* 34.48, pp. 15619–15623. ISSN: 0006-2960. DOI: 10.1021/bi00048a003.

Than, Manuel E. et al. (2004). "The 2.0Å crystal structure and substrate specificity of the KDEL-tailed cysteine endopeptidase functioning in programmed cell death of Ricinus communis endosperm". en. In: *Journal of Molecular Biology* 336.5, pp. 1103–1116. ISSN: 0022-2836. DOI: 10.1016/j.jmb.2003.12.075.

Thatcher, Louise F. et al. (2012). "A highly conserved effector in Fusarium oxysporum is required for full virulence on Arabidopsis". eng. In: *Molecular plant-microbe interactions: MPMI* 25.2, pp. 180–190. ISSN: 0894-0282. DOI: 10.1094/MPMI-08-11-0212.

The Potato Genome Sequencing Consortium (2011). "Genome sequence and analysis of the tuber crop potato". en. In: *Nature* 475.7355, pp. 189–195. ISSN: 1476-4687. DOI: 10.1038/nature10158.

Thomma, Bart P. H. J. et al. (1998). "Separate jasmonate-dependent and salicylate-dependent defense-response pathways in Arabidopsis are essential for resistance to distinct microbial pathogens". In: *Proceedings of the National Academy of Sciences* 95.25, pp. 15107–15111. DOI: 10.1073/pnas.95.25.15107.

Thompson, John N. and Jeremy J. Burdon (1992). "Gene-for-gene coevolution between plants and parasites". en. In: *Nature* 360.6400, pp. 121–125. ISSN: 1476-4687. DOI: 10.1038/360121a0.

Tian, Miaoying, Brett Benedetti, and Sophien Kamoun (2005). "A second kazal-like protease inhibitor from Phytophthora infestans inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato". en. In: *Plant Physiology* 138.3, pp. 1785–1793. ISSN: 0032-0889, 1532-2548. DOI: 10.1104/pp.105.061226.

Tian, Miaoying and Sophien Kamoun (2005). "A two disulfide bridge Kazal domain from Phytophthora exhibits stable inhibitory activity against serine proteases of the subtilisin family". In: *BMC Biochemistry* 6, p. 15. ISSN: 1471-2091. DOI: 10.1186/1471-2091-6-15.

Tian, Miaoying et al. (2004). "A kazal-like extracellular serine protease inhibitor from phytophthora infestans targets the tomato pathogenesis-related protease P69B". en. In: *Journal of Biological Chemistry* 279.25, pp. 26370–26377. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M400941200.

Tian, Miaoying et al. (2007). "A Phytophthora infestans cystatin-like protein targets a novel tomato papain-like apoplastic protease". en. In: *Plant Physiology* 143.1, pp. 364–377. ISSN: 0032-0889, 1532-2548. DOI: 10.1104/pp.106.090050.

Tognolli, Michael et al. (2002). "Analysis and expression of the class III peroxidase large gene family in Arabidopsis thaliana". eng. In: *Gene* 288.1-2, pp. 129–138. ISSN: 0378-1119. DOI: 10.1016/s0378-1119(02)00465-1.

Tornero, P., V. Conejero, and P. Vera (1997). "Identification of a new pathogen-induced member of the subtilisin-like processing protease family from plants". eng. In: *The Journal of Biological Chemistry* 272.22, pp. 14412–14419. ISSN: 0021-9258. DOI: doi:10.1074/jbc.272.22.14412.

Tornero, P. et al. (1996). "Characterization of LRP, a leucine-rich repeat (LRR) protein from tomato plants that is processed during pathogenesis". eng. In: *The Plant Journal:*

For Cell and Molecular Biology 10.2, pp. 315–330. ISSN: 0960-7412. DOI: 10.1046/j.1365-313x.1996.10020315.x.

Tripathi, Diwaker, Yu-Lin Jiang, and Dhirendra Kumar (2010). "SABP2, a methyl salicylate esterase is required for the systemic acquired resistance induced by acibenzolar-S-methyl in plants". en. In: *FEBS Letters* 584.15, pp. 3458–3463. ISSN: 1873-3468. DOI: 10.1016/j.febslet.2010.06.046.

Trudel, Jean et al. (1998). "Several thaumatin-like proteins bind to -1,3-glucans". In: *Plant Physiology* 118.4, pp. 1431–1438. ISSN: 0032-0889. DOI: 10.1104/pp.118.4.1431.

Tsaban, Tomer et al. (2022). "Harnessing protein folding neural networks for peptide–protein docking". en. In: *Nature Communications* 13.1, p. 176. ISSN: 2041-1723. DOI: 10.1038/s41467-021-27838-9.

Uknes, S et al. (1992). "Acquired resistance in Arabidopsis." In: *The Plant Cell* 4.6, pp. 645–656. ISSN: 1040-4651. DOI: 10.1105/tpc.4.6.645.

Vallée, B.s. et al. (2003). "Peptides corresponding to the N- and C-terminal parts of PEBP are well-structured in solution: new insights into their possible interaction with partners in vivo". en. In: *The Journal of Peptide Research* 61.2, pp. 47–57. ISSN: 1399-3011. DOI: 10.1034/j.1399-3011.2003.21039.x.

Van Dijk, Michiel et al. (2021). "A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050". en. In: *Nature Food* 2.7, pp. 494–501. ISSN: 2662-1355. DOI: 10.1038/s43016-021-00322-9.

Van Esse, H. Peter et al. (2008). "The Cladosporium fulvum virulence protein Avr2 inhibits host proteases required for basal defense". In: *The Plant Cell* 20.7, pp. 1948–1963. ISSN: 1040-4651. DOI: 10.1105/tpc.108.059394.

Van loon, L. C. (2000). "Systemic induced resistance". In: *Mechanisms of Resistance to Plant Diseases*. Springer, pp. 521–575. ISBN: 978-94-011-3937-3.

Van Loon, L.C., M. Rep, and C.M.J. Pieterse (2006). "Significance of inducible defense-related proteins in infected plants". In: *Annual Review of Phytopathology* 44.1, pp. 135–162. DOI: 10.1146/annurev.phyto.44.070505.143425.

Van den Burg, Harrold A. et al. (2006). "Cladosporium fulvum Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection". eng. In: *Molecular plant-microbe interactions: MPMI* 19.12, pp. 1420–1430. ISSN: 0894-0282. DOI: 10.1094/MPMI-19-1420.

Van der Burgh, Aranka M. and Matthieu H. A. J. Joosten (2019). "Plant immunity: thinking outside and inside the box". eng. In: *Trends in Plant Science* 24.7, pp. 587–601. ISSN: 1878-4372. DOI: 10.1016/j.tplants.2019.04.009.

Van der Does, H. Charlotte et al. (2008). "The presence of a virulence locus discriminates Fusarium oxysporum isolates causing tomato wilt from other isolates". eng. In: *Environmental Microbiology* 10.6, pp. 1475–1485. ISSN: 1462-2920. DOI: 10.1111/j.1462-2920.2007.01561.x.

Van der Wel, H. and K. Loeve (1972). "Isolation and characterization of thaumatin I and II, the sweet-tasting proteins from Thaumatococcus daniellii Benth". eng. In: *European*

Journal of Biochemistry 31.2, pp. 221–225. ISSN: 0014-2956. DOI: 10.1111/j.1432-1033.1972.tb02522.x.

Van Kan, Jan A. L. et al. (2017). "A gapless genome sequence of the fungus Botrytis cinerea". eng. In: *Molecular Plant Pathology* 18.1, pp. 75–89. ISSN: 1364-3703. DOI: 10.1111/mpp.12384.

Van Loon, L. C. and J. F. Antoniw (1982). "Comparison of the effects of salicylic acid and ethephon with virus-induced hypersensitivity and acquired resistance in tobacco". en. In: *Netherlands Journal of Plant Pathology* 88.6, pp. 237–256. ISSN: 1573-8469. DOI: 10.1007/BF02000130.

Van Loon, L. C. and A. Van Kammen (1970). "Polyacrylamide disc electrophoresis of the soluble leaf proteins from Nicotiana tabacum var. 'Samsun' and 'Samsun NN': II. Changes in protein constitution after infection with tobacco mosaic virus". en. In: *Virology* 40.2, pp. 199–211. ISSN: 0042-6822. DOI: 10.1016/0042-6822(70)90395-8.

Verschaffelt, Pieter et al. (2021). "Unipept desktop: a faster, more powerful metaproteomics results analysis tool". In: *Journal of Proteome Research* 20.4, pp. 2005–2009. ISSN: 1535-3893. DOI: 10.1021/acs.jproteome.0c00855.

Vlot, A. Corina, D'Maris Amick Dempsey, and Daniel F. Klessig (2009). "Salicylic Acid, a multifaceted hormone to combat disease". eng. In: *Annual Review of Phytopathology* 47, pp. 177–206. ISSN: 0066-4286. DOI: 10.1146/annurev.phyto.050908.135202.

Vogel, Christine and Edward M Marcotte (2009). "Absolute abundance for the masses". In: *Nature biotechnology* 27.9, pp. 825–826. ISSN: 1087-0156. DOI: 10.1038/nbt0909-825.

Wally, O. and Z. K. Punja (2010). "Enhanced disease resistance in transgenic carrot (Daucus carota L.) plants over-expressing a rice cationic peroxidase". eng. In: *Planta* 232.5, pp. 1229–1239. ISSN: 1432-2048. DOI: 10.1007/s00425-010-1252-4.

Wan, Jiangxue et al. (2021). "Cell wall associated immunity in plants". en. In: *Stress Biology* 1.1, p. 3. ISSN: 2731-0450. DOI: 10.1007/s44154-021-00003-4.

Wang, Andrea L. et al. (2020a). "Comparison of glycoprotein separation reveals greater impact of carbohydrates and disulfides on electrophoretic mobility for CE-SDS versus SDS-PAGE". en. In: *Journal of Pharmaceutical and Biomedical Analysis* 180, p. 113006. ISSN: 0731-7085. DOI: 10.1016/j.jpba.2019.113006.

Wang, Peng et al. (2020b). "The genome evolution and domestication of tropical fruit mango". In: *Genome Biology* 21.1, p. 60. ISSN: 1474-760X. DOI: 10.1186/s13059-020-01959-8.

Wang, Shuaishuai et al. (2021a). "Cleavage of a pathogen apoplastic protein by plant subtilases activates host immunity". eng. In: *The New Phytologist* 229.6, pp. 3424–3439. ISSN: 1469-8137. DOI: 10.1111/nph.17120.

Wang, Ying et al. (2021b). "Regulation and function of defense-related callose deposition in plants". en. In: *International Journal of Molecular Sciences* 22.5, p. 2393. ISSN: 1422-0067. DOI: 10.3390/ijms22052393.

Wang, Yu et al. (2015). "Systematic analysis of maize class III peroxidase gene family reveals a conserved subfamily involved in abiotic stress response". en. In: *Gene* 566.1, pp. 95–108. ISSN: 0378-1119. DOI: 10.1016/j.gene.2015.04.041.

Waterhouse, Andrew et al. (2018a). "SWISS-MODEL: homology modelling of protein structures and complexes". en. In: *Nucleic Acids Research* 46.W1, W296–W303. ISSN: 0305-1048. DOI: 10.1093/nar/gky427.

Waterhouse, Robert M et al. (2018b). "BUSCO applications from quality assessments to gene prediction and phylogenomics". In: *Molecular Biology and Evolution* 35.3, pp. 543–548. ISSN: 0737-4038. DOI: 10.1093/molbev/msx319.

Weber, Ernst et al. (2011). "A modular cloning system for standardized assembly of multigene constructs". eng. In: *PloS One* 6.2, e16765. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0016765.

Wei, Chia-Fong et al. (2007). "A Pseudomonas syringae pv. tomato DC3000 mutant lacking the type III effector HopQ1-1 is able to cause disease in the model plant Nicotiana benthamiana". eng. In: *The Plant Journal: For Cell and Molecular Biology* 51.1, pp. 32–46. ISSN: 0960-7412. DOI: 10.1111/j.1365-313X.2007.03126.x.

White, R. F. (1979). "Acetylsalicylic acid (aspirin) induces resistance to tobacco mosaic virus in tobacco". en. In: *Virology* 99.2, pp. 410–412. ISSN: 0042-6822. DOI: 10.1016/0042-6822(79)90019-9.

Wiermer, Marcel, Bart J Feys, and Jane E Parker (2005). "Plant immunity: the EDS1 regulatory node". en. In: *Current Opinion in Plant Biology*. Biotic interactions 8.4, pp. 383–389. ISSN: 1369-5266. DOI: 10.1016/j.pbi.2005.05.010.

Winsor, Geoffrey L. et al. (2016). "Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database". eng. In: *Nucleic Acids Research* 44.D1, pp. D646–653. ISSN: 1362-4962. DOI: 10.1093/nar/gkv1227.

Witzel, Katja et al. (2011). "Comparative evaluation of extraction methods for apoplastic proteins from maize leaves". In: *Plant Methods* 7, p. 48. ISSN: 1746-4811. DOI: 10.1186/1746-4811-7-48.

Wiśniewski, Jacek R. et al. (2012). "Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma". eng. In: *Molecular Systems Biology* 8, p. 611. ISSN: 1744-4292. DOI: 10.1038/msb.2012.44.

Wu, Yue et al. (2012). "The Arabidopsis NPR1 protein is a receptor for the plant defense hormone salicylic acid". In: *Cell Reports* 1.6, pp. 639–647. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2012.05.008.

Xia, Yiji et al. (2004). "An extracellular aspartic protease functions in Arabidopsis disease resistance signaling". In: *The EMBO Journal* 23.4, pp. 980–988. ISSN: 0261-4189. DOI: 10.1038/sj.emboj.7600086.

Xie, Weilong and Paul H. Goodwin (2009). "A PRp27 gene of Nicotiana benthamiana contributes to resistance to Pseudomonas syringae pv. tabaci but not to Colletotrichum destructivum or Colletotrichum orbiculare". en. In: *Functional Plant Biology* 36.4, pp. 351–361. ISSN: 1445-4416. DOI: 10.1071/FP08241.

Xu, Xiaoyang et al. (2016). "Genome-wide characterization of the -1,3-glucanase gene family in Gossypium by comparative analysis". en. In: *Scientific Reports* 6.1, p. 29044. ISSN: 2045-2322. DOI: 10.1038/srep29044.

Yaish, Mahmoud W.F. et al. (2006). "Cold-active winter rye glucanases with ice-binding capacity". In: *Plant Physiology* 141.4, pp. 1459–1472. ISSN: 0032-0889. DOI: 10.1104/pp.106.081935.

Yang, Fen et al. (2015). "Unraveling incompatibility between wheat and the fungal pathogen Zymoseptoria tritici through apoplastic proteomics". In: *BMC Genomics* 16.1, p. 362. ISSN: 1471-2164. DOI: 10.1186/s12864-015-1549-6.

Yang, Zhaohe et al. (2019). "Identification and characterization of the PEBP family genes in moso bamboo (Phyllostachys heterocycla)". en. In: *Scientific Reports* 9.1, p. 14998. ISSN: 2045-2322. DOI: 10.1038/s41598-019-51278-7.

Yang, Ziheng, Wendy S. W. Wong, and Rasmus Nielsen (2005). "Bayes empirical bayes inference of amino acid sites under positive selection". en. In: *Molecular Biology and Evolution* 22.4, pp. 1107–1118. ISSN: 0737-4038. DOI: 10.1093/molbev/msi097.

Yang, Ziheng et al. (2000). "Codon-substitution models for heterogeneous selection pressure at amino acid sites". en. In: *Genetics* 155.1, pp. 431–449. ISSN: 0016-6731, 1943-2631. DOI: https://doi.org/10.1093/genetics/155.1.431.

Yeung, Kam C. et al. (2001). "Raf kinase inhibitor protein interacts with NF-B-inducing kinase and TAK1 and inhibits NF-B activation". In: *Molecular and Cellular Biology* 21.21, pp. 7207–7217. DOI: 10.1128/MCB.21.21.7207-7217.2001.

Yoshino, Kae et al. (2012). "Cell death of Nicotiana benthamiana is induced by secreted protein NIS1 of Colletotrichum orbiculare and is suppressed by a homologue of CgDN3". eng. In: *Molecular plant-microbe interactions: MPMI* 25.5, pp. 625–636. ISSN: 0894-0282. DOI: 10.1094/MPMI-12-11-0316.

Yu, Xiao-Ming and Marilyn Griffith (1999). "Antifreeze proteins in winter rye leaves form oligomeric complexes". In: *Plant Physiology* 119.4, pp. 1361–1370. ISSN: 0032-0889. DOI: 10.1104/pp.119.4.1361.

Yuan, Minhang et al. (2021). "Pattern-recognition receptors are required for NLR-mediated plant immunity". en. In: *Nature* 592.7852, pp. 105–109. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03316-6.

Zhang, Yang and Jeffrey Skolnick (2004). "Scoring function for automated assessment of protein structure template quality". eng. In: *Proteins* 57.4, pp. 702–710. ISSN: 1097-0134. DOI: 10.1002/prot.20264.

Zhang, Yuelin and Xin Li (2019). "Salicylic acid: biosynthesis, perception, and contributions to plant immunity". en. In: *Current Opinion in Plant Biology*. Biotic interactions 50, pp. 29–36. ISSN: 1369-5266. DOI: 10.1016/j.pbi.2019.02.004.

Zhao, Min et al. (2018). "An integrated analysis of mRNA and sRNA transcriptional profiles in tomato root: Insights on tomato wilt disease". In: *PLoS ONE* 13.11, e0206765. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0206765.