Evolution of an extracellular pathogen recognition mechanism in Solanaceae



Jiorgos Kourelis

The King's Hall and College of Brasenose

Department of Plant Sciences

University of Oxford

Supervised by Professor Renier A. L. van der Hoorn

Thesis submitted for the degree of Doctor of Philosophy

Trinity 2018

Evolution of an extracellular pathogen recognition mechanism in Solanaceae

Jiorgos Kourelis

The King's Hall and College of Brasenose

Department of Plant Sciences

University of Oxford

Supervised by Professor Renier A.L. van der Hoorn

Thesis submitted for the degree of Doctor of Philosophy

Trinity 2018

Abstract

The innate immune system of plants depends on resistance (*R*) genes which provide durable resistance to pathogens in nature. Although many *R* genes have been identified, the molecular mechanism by which they function is only understood for a few.

To study these mechanisms, I present four papers on the molecular mechanisms by which *R* gene products function, and one paper on a resource useful for the plant-pathogen community. First, I present a review and meta-analysis of the molecular mechanisms by which *R* gene products function, thereby identifying nine molecular mechanisms conferring pathogen resistance in plants. Second, I describe the evolutionary processes which have given rise to an indirect apoplastic recognition mechanism in the nightshade (Solanaceae) family. Tomato varieties carrying *Cf-2*, which encodes for a receptor-like protein, recognize the fungal-derived small secreted protein Avr2. This recognition requires the tomato gene *Rcr3*, which encodes for a papain-like cysteine protease. I show that *Cf-2* has likely evolved very recently, implicating that molecular recognition is a novel molecular function of the pre-existing *Rcr3*.

I discuss this in a broader evolutionary context of similar recognition mechanisms. Third, I comment on a paper in which a similar resistance mechanism was engineered to alter its recognition spectrum. Fourth, I identify the amino acid residues in Rcr3 required for *Cf-2*-dependent recognition of Avr2 and I engineer another protease to replace Rcr3 for the recognition of Avr2. This potentially enables future engineering of this molecular mechanism for an altered recognition spectrum. Finally, using the proteome derived from re-annotated gene models of the plant model organism *Nicotiana benthamiana*, I show that the apoplast is enriched for enzymes with hydrolytic functions.

By elucidating the evolution of a pathogen recognition mechanism in Solanaceae and establishing the engineering of this mechanism, this thesis provides a framework for the generation of an altered recognition spectrum not found in nature.

Acknowledgements

I am grateful to many people who have made this DPhil possible, and to those who have inspired me before in one way or another to pursue a DPhil.

I would like to thank Renier van der Hoorn for offering me the DPhil position in Oxford and his supervision. I am thankful for the occasional push to finish projects (e.g. the review) as well as the academic freedom in the research allowing me to occasionally try out one of my many ideas. I also greatly appreciated being asked to comment on papers and grants.

I am lucky to have supervised several very talented students in the lab: Ana Bravo, Samantha Royston, Ruby O'Grady, Parvinderdeep Kahlon, Oliver Mattinson, and Sonja Krauter. Your contributions were critical for many parts of this thesis.

I have had many collaborations and discussions with lab members and people from the department, all of which have contributed to this thesis. Thank you to Friederike for the many discussions on pretty much any topic, we disagreed most of the time which just made the conversation more fun. To Marcel for the discussions on Catalan independence, Judith for the collaboration and the exchange of materials, and both Marcel and Judith for all the parties and fun. To Shivani for providing me with Avr2 protein, to Barney and Dana for the plasmid-sharing, to Tee for all the help with R, to Emma for being wrong in the pronunciation of the letter aitch but nevertheless persisting, to Bala and Tram for being the first to arrive in Oxford and for setting up the lab there, to Phillipe for interesting scientific ideas, to Kyoko for all the collaborations, to Felix for help with coding, and Marcel, Felix, and Laura for “Yoining” to the “Yym” occasionally (and always late at great dismay of Felix). To Maria for the dinners at her place, to Pierre for the French conversation (Ça va? Ça va, ça va? Ça va!). To mama Sueldo (Daniela) for insights on the Netherlands, to Wanda - also known as Flo - for inspiring us with her athletic career, to Riccardo for sorting agrichemicals the Italian way, and all others.

I would like to thank Urszula, Sarah, and Caroline for excellent technical support, and our collaborators Pitter Huesgen and Fatih Demir in Juelich, and Farnusch Kaschani in Essen.

For my funding I thank The Clarendon fund, which funded most of my DPhil together with Brasenose College, and to Brasenose for funding most of my conference trips.

Many people have inspired me in one way or another to pursue a DPhil. During my BSc and MSc I was lucky to encounter many professors as well as PhDs and post-docs with a passion for science. From Amsterdam I would like to thank Christa Testerink, Carlos Galvan-Ampudia, Magdalena Julkowska, Dorota Kawa, Teun Munnik, Ringo van Wijk, Frank Takken, Rob Schuurink, and Michel Haring. From my time in Riverside I would like to thank Sean Cutler and Assaf Mosquna. Finally, I am very thankful to Nikos Vlassis, for many conversations and games of Carcassonne which truly encouraged an academic vision of life. I am also thankful to Hans Smit, for the first encounters with the subject of biology, and to my aunt Helen Geertman for providing a place for me to stay - with a full, albeit gluten-free, fridge - during my BSc and whenever I am back in Amsterdam.

Finally, I would like to thank my family (my brother Stefanos Kourelis, my mom Janine van der Kolk-Geertman and her husband Dick van der Kolk (and his children Leonie and Pim van der Kolk), my father Taxiarchis Kourelis and his wife Dafni Alverti with my youngest brother Inachos Kourelis, Wietske Smit, my grandparents Hans-Bastiaan Geertman, Anna Gerda Geertman-Hoepman, Georgios Kourelis, and Athanassia Kourelis, and my uncle Spyridon Kourelis and aunt Myrsini Alverti) and friends for their support, and Evie, for putting up with me.

Thank you all,



Jiorgos Kourelis

Table of Contents

Chapter 1: General introduction 1

Resistance to pathogens is determined by resistance (*R*) genes 3

Cf genes encode for receptor-like proteins conferring resistance to *C. fulvum* in a gene-for-gene relationship 3

The molecular mechanisms by which *Cf* genes confer recognition of effectors are not well understood 4

The molecular mechanisms by which proteases contribute to basal resistance are not well understood 6

Protease-inhibitor interactions in the host apoplast 7

The aim of this thesis is to characterize the evolution and mechanism of a pathogen recognition mechanism 9

Chapter 2: Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function 13

Introduction 16

Overview and occurrence of the nine mechanisms 19

Subclade Ia: Extracellular Perception 20

Mechanism 1: Direct Perception at the Cell Surface 20

Mechanism 2: Indirect Extracellular Effector Perception 23

Subclade Ib: Intracellular Perception 25

Mechanism 3: Direct Intracellular Effector Recognition—Trigger-Happy NLRs 25

Mechanism 4: Indirect Intracellular Recognition—Decoys and Guardees 27

Mechanism 5: NLR-IDs—Sensors Embedded within Receptors 31

Mechanism 6: Executor Genes—Promoter Traps for TAL Effectors 33

Clade II: Loss of Susceptibility 34

Mechanism 7: Active Loss of Susceptibility—Blocking Key Pathogen Strategies 34

Mechanism 8: Passive Loss of Susceptibility—Losing Interaction with Key Host Targets 35

Mechanism 9: Passive Loss of Susceptibility by Host Reprogramming 36

Discussion 38

Supplemental Data 40

Chapter 3: Evolution of a guarded decoy protease and its receptor in Solanaceae 41

Introduction: 44

Results: 46
Co- expression of Cf-2, Rcr3, and Avr2 triggers a robust HR in *N. benthamiana* 46
N. benthamiana Rcr3 is a pseudogene and can be resurrected, but it cannot participate in Cf-2-dependent Avr2 recognition 48
Pip1: an important immune protease in tomato 52
Rcr3 is conserved across the Solanaceae family 55
Three residues determine PLCP sensitivity for Avr2 58
Cf-2 belongs to a new-world specific clade of the Hcr2 gene family 60
Discussion: 64
Material & Methods: 66
Supplemental figures 77
Chapter 4: Decoy Engineering: The Next Step in Resistance Breeding 85
Chapter 5: Induction of the hypersensitive response by Rcr3 resides in few Rcr3-specific amino acid residues 93
Introduction 96
Results 97
Distant Rcr3 homologs can be engineered for Cf-2-dependent Avr2 recognition 97
A screen of Rcr3/Pip1 hybrids reveals three distinct regions of Rcr3 required for Cf-2-dependent Avr2 recognition 100
Mutagenesis in Rcr3/Pip1 hybrid proteases pinpoints 12 residues required for Cf-2/Avr2-dependent HR 103
Rcr3 and Pip1 have distinct substrate specificity 106
Discussion 107
Material & Methods 109
Supplemental figures 120
Chapter 6: Re-annotated *Nicotiana benthamiana* gene models for enhanced proteomics and reverse genetics 125
Introduction 128
Results and Discussion 130
Re-annotation of gene-models in the *N. benthamiana* genome assemblies 130
The new proteome database is more complete, more sensitive and accurate, and relatively small 131
Improved annotation of the apoplastic proteome of *N. benthamiana* 135
Pseudogenization in the subtilisin family is consistent with a contracting functional genome 138
Improved accuracy for genome editing: the subtilase gene-family 140
Material & Methods 142

Supplemental figures 150
Chapter 7: General Discussion and conclusion 155
The key messages from this thesis 157
The molecular mechanisms of *R* genes 157
Understanding the evolution of *R* genes may shed light on their mechanisms 162
Decoy swapping as a strategy to engineer novel recognition 164
The *N. benthamiana* apoplastic proteome 164
Conclusions 165
Appendix 167
Bodyguards: Pathogen-Derived Decoys That Protect Virulence Factors 169
References 175

Chapter 1

General Introduction

Resistance to pathogens is determined by resistance (R) genes

Plant resistance to pathogens is determined by resistance (*R*) genes. These *R* genes often encode for receptors that recognize pathogen-associated molecular patterns (PAMPs) and effectors. Identifying *R* genes has been a main focus in plant-pathogen research, and has led to the discovery of the molecular identity of approximately 300 *R* genes (Kourelis and van der Hoorn, 2018).

One plant-pathogen which has served as a model for plant-microbe interactions is the tomato leaf mould pathogen *Cladosporium fulvum* (syn. *Passalora fulva*). *C. fulvum* is a biotrophic fungal pathogen of cultivated and wild tomato species (genus *Solanum*, subgenus *Solanum sensu stricto*, section *Lycopersicon*). *C. fulvum* does not produce haustoria - structures that are produced by other pathogens and used to invade host cells - so its dialogue with the host occurs completely in the leaf extracellular space (apoplast) (Thomma *et al.*, 2005). To create a favourable environment *C. fulvum* secretes a wide range of effectors in the apoplast. These effectors are small molecules and proteins which alter host-cell structure and function, thereby facilitating infection (Hogenhout *et al.*, 2009). Tomato leaf mould was a disease of high economic importance up until the introduction of tomato cultivars in which *R* genes were introgressed from wild tomato species which confer resistance to *C. fulvum* (de Wit *et al.*, 2012). These *R* genes are called *Cf* genes and confer recognition of *C. fulvum* races secreting specific effectors in a gene-for-gene relationship. Recognition of their cognate ligand triggers the hypersensitive response (HR), a form of localized programmed cell-death.

Cf genes encode for receptor-like proteins conferring resistance to C. fulvum in a gene-for-gene relationship

All cloned *Cf* genes encode for receptor-like proteins (RLPs) with extracellular leucine-rich repeat (LRR) domains. Cloned *Cf* genes belong to two gene-families: the *Hcr9* (for homologous to *Cladosporium* resistance gene *Cf-9*) gene-family residing on chromosome 1

and the *Hcr2* (for homologous to *Cladosporium* resistance gene *Cf-2*) gene family residing on chromosome 6.

Six *Cf* genes have been cloned from the *Hcr9* gene-family. *Cf-9*, which originates from the currant tomato (*Solanum pimpinellifolium*), was the first cloned *R* gene encoding a RLP (Jones *et al.*, 1994). *Cf-9* confers recognition of a small cysteine-rich *C. fulvum* effector called Avr9 (van Kan *et al.*, 1991). Similar to *Cf-9* is *9DC*, which also confers recognition of Avr9 and is also present in currant tomato (van der Hoorn *et al.*, 2001a). By contrast, *Cf-4*, which was introgressed from *Solanum habrochaites*, confers recognition of Avr4 (Thomas *et al.*, 1997; Joosten *et al.*, 1994). *Hcr9-4E* is present in the same haplotype as *Cf-4* and confers recognition of Avr4E (Takken *et al.*, 1999; Westerink *et al.*, 2004). Finally, *Cf-19*, present in cultivated tomato, and *Hcr9-9B*, present on the same haplotype as *Cf-9*, recognize as-of-yet unidentified *C. fulvum* effectors (Zhao *et al.*, 2016; Chakrabarti *et al.*, 2009).

Two *Cf* genes have been isolated from the Hcr2 locus. *Cf-2* was introgressed from the currant tomato (Dixon *et al.*, 1996), while *Cf-5* comes from the cherry tomato (*S. lycopersicum* var. *cerasiforme*) (Seear and Dixon, 2003). *Cf-2* confers recognition of Avr2 (Luderer *et al.*, 2002), while *Cf-5* confers recognition of Avr5 (Mesarich *et al.*, 2014).

The molecular mechanisms by which Cf genes confer recognition of effectors are not well understood

How these *Cf* genes confer recognition of these *C. fulvum* effectors is not known, and the molecular roles of only a few of these effectors have been shown. Through domains swaps between Cf-4 and Cf-9, the residues required for differential effector recognition between these proteins have been identified (van der Hoorn *et al.*, 2001b; Wulff *et al.*, 2001). This has also been done for Cf-9 and Hcr9-9B (Chakrabarti *et al.*, 2009). It is tempting to speculate that the products of these *Cf* genes interact directly with their cognate ligands. However, no direct

interaction could be found between Cf-9 and Avr9 (Luderer *et al.*, 2001). Avr9 does bind to a unknown component present in the plasma membrane of Solanaceae species tentatively called the high-affinity binding site (HABS) (Kooman-Gersmann *et al.*, 1996). The interaction of Avr9 with the HABS correlates with its capacity to induce *Cf-9*-dependent HR (Kooman-Gersmann *et al.*, 1998). This indicates a possible indirect recognition mechanism. The recent elucidation of the structure of Avr4 confirms it is a chitin-binding protein, and this chitin-binding capacity is not required for the induction of *Cf-4*-dependent HR (Hurlburt *et al.*, 2018). Instead, these studies showed the importance of Avr4's capacity to resist proteolysis in the plant apoplast for recognition by *Cf-4* (Hurlburt *et al.*, 2018).

Of the cloned *Cf* genes, only *Cf-2* has a proposed molecular mechanism of recognition. *Cf-2* triggers an autonecrotic response in the presence of a genetically unlinked gene named *Ne* (for necrosis) (Langford, 1948). Later it was found that *Cf-2* requires *Rcr3* for its function (Dixon *et al.*, 2000). *Rcr3* encodes for a papain-like cysteine protease (PLCP), and the cultivated tomato *Rcr3* homolog ($Rcr3^{lyc}$) was identified as the necrosis-inducing *Ne* gene (Krüger *et al.*, 2002). Co-introgression of the currant tomato *Rcr3* homolog ($Rcr3^{pim}$ or *ne*) is required for the suppression of this autonecrosis (Krüger *et al.*, 2002). Avr2 act as a protease inhibitor, inhibiting Rcr3 (Rooney *et al.*, 2005). It is the interaction of Avr2 with Rcr3 which is recognized by *Cf-2* (Rooney *et al.*, 2005). *Rcr3* does not seem to contribute to basal resistance to *C. fulvum* in the absence of *Cf-2* (Ilyas *et al.*, 2015). By contrast, Avr2 does contribute to the virulence of *C. fulvum* in the absence of *Cf-2* (van Esse *et al.*, 2008). This has led to the hypothesis that Rcr3 acts as a decoy, trapping *C. fulvum* in a recognition event (van der Hoorn and Kamoun, 2008). Indeed, Avr2 also inhibits other proteases, including the Rcr3-like Pip1 (Shabab *et al.*, 2008; van Esse *et al.*, 2008). *Pip1* does contribute to basal resistance to *C. fulvum* in the absence of *Cf-2*, unlike *Rcr3*, and has therefore been proposed to be the operative target of Avr2 (Ilyas *et al.*, 2015).

The molecular mechanisms by which proteases contribute to basal resistance are not well understood

Pip1 is an important general immune protease which contributes to resistance against diverse pathogens including *C. fulvum*, *Phytophthora infestans*, and *Pseudomonas syringae* pv. *tomato* DC3000 (Ilyas *et al.*, 2015). Rcr3 only has a minor role in basal resistance against *P. infestans* (Song *et al.*, 2009; Ilyas *et al.*, 2015). Other unrelated proteases have also been implicated in basal resistance against pathogens. For example, the *Nicotiana benthamiana* protease *NbC14* was found to contribute to resistance against *P. infestans* (Kaschani *et al.*, 2010). However, the cellular pathways through which host proteases contribute to immunity, and conversely how effectors which inhibit them contribute to pathogen virulence, are not fully understood. Three mechanisms by which apoplastic proteases may contribute to resistance against pathogens have been identified or proposed: i) release of immunogenic elicitors by proteolytic activity, ii) activation/inactivation of host and/or pathogen substrates involved in resistance/susceptibility or virulence, respectively, or iii) as atypical co-receptors like Rcr3.

i) Proteases are likely involved in the maturation of host-derived danger-associated molecular patterns (DAMPs) in the apoplast. In maize (*Zea mays*), PLCPs are required to release the bioactive *Z. mays* immune signalling peptide 1 (Zip1) from its precursor protein (Ziemann *et al.*, 2018). Zip1 enhances resistance to the biotrophic fungus *Ustilago maydis* (Ziemann *et al.*, 2018). In Arabidopsis (*Arabidopsis thaliana*), several subtilisin-like serine proteases as well as a type-II metacaspase release bioactive peptides from diverse precursor proteins (Srivastava *et al.*, 2008; Srivastava *et al.*, 2009; Engineer *et al.*, 2014; Ghorbani *et al.*, 2016; Schardon *et al.*, 2016; Wrzaczek *et al.*, 2015). Also the release of recognition-ready epitopes of pathogen-associated molecular patterns (PAMPs) likely requires the activity of proteases and other hydrolases. For example, the release of immunogenic fragments of pathogen-derived chitin in Arabidopsis requires the activity of the metazoan lysozyme-like hydrolase LYS1 (Liu *et al.*,

2014). Additionally, the release of immunogenic flagellin epitopes is likely due to the concerted activity of various hydrolases (Felix *et al.*, 1999; Schiermeyer *et al.*, 2009) (Buscaill et al., submitted). No plant protease has yet been demonstrated to be directly involved in the release of immunogenic peptides from PAMPs.

ii) Alternatively, proteases may be involved in activating or inactivating proteins or metabolites involved in the apoplast during host-pathogen interactions. It is surprising that there are no convincing examples of apoplastic proteases involved in such a role in host-pathogen interactions. The closest example is that of the tomato matrix metalloproteases Sl2-MMP and Sl3-MMP, which can cleave and inactivate the pathogenesis-related serine protease P69B (Zimmermann *et al.*, 2016). Whether this is functionally important in host-pathogen interactions is not known. Additionally, P69C proteolytically processes the tomato LRR-containing protein LRP in a pathogenic context (Tornero *et al.*, 1996).

iii) Finally, apoplastic proteases may act as ligand-binding domains involved in molecular recognition mediated by cell-surface receptors. It is possible that in addition to Rcr3 other proteases are also required for the recognition of pathogen-derived components through such a mechanism. For example, introduction of the *Hwi1* locus from wild rice (*Oryza rufipogon*) in *indica* rice (*Oryza sativa indica*) causes a hybrid incompatibility dependent on the rare *Hwi2* locus (Chen *et al.*, 2014). *Hwi1* encodes for two LRR containing receptor-like kinases (RLKs) called 25L1 and 25L2, and *Hwi2* encodes for a subtilisin-like serine protease which differs by only one amino acid from the compatible allele (Chen *et al.*, 2014). It is tempting to speculate that this subtilisin-like serine protease acts in a similar manner to Rcr3.

Protease-inhibitor interactions in the host apoplast

Apoplastic proteases therefore appear to have a general role in basal resistance against pathogens. *C. fulvum* is not the only pathogen which secretes protease inhibitors. In fact,

proteases seem to be important virulence "hubs" on which the effectors of diverse pathogens converge, as was described previously for other intracellular effector targets (Mukhtar *et al.*, 2011).

P. infestans, for example, secretes the distinct and structurally divergent Kazal-like serine protease inhibitors *Pi*EPI1 and *Pi*EPI10 which inhibit the pathogenesis-related serine protease P69B from tomato (Tian *et al.*, 2004; Tian *et al.*, 2005). The expression of these effector genes is increased during the infection of tomato, indicative of a contribution to virulence (Tian *et al.*, 2004; Tian *et al.*, 2005). Proteases are also important virulence targets in monocots, as demonstrated by the *U. maydis* effector Pit2 which is required for virulence and inhibits maize PLCPs (Mueller *et al.*, 2013).

Rcr3 is also inhibited by several other pathogen-derived protease inhibitors. The cystatin-like inhibitors *Pi*EPIC1 and *Pi*EPIC2B from *P. infestans* can inhibit the PLCPs Rcr3 (Song *et al.*, 2009), Pip1 (Tian *et al.*, 2007), and C14 (Kaschani *et al.*, 2010). This does not, however, trigger a *Cf-2*-dependent recognition of these effectors (Song *et al.*, 2009). These effectors and other similar effectors appear to be important contributors to virulence on diverse hosts. A single substitution, Q111R, in *Phytophthora mirabilis Pm*EPIC1 has been proposed to underlie biochemical specialization of this effector to protease targets of *Mirabilis jalapa* following a host jump (Dong *et al.*, 2014). Furthermore, the *Phytophthora palmivora*-specific cystatin-like effector PpalEPIC8 inhibits papaya (*Carica papaya*) PLCPs and is required for virulence (Gumtow *et al.*, 2017). Bacterial pathogens, in addition to fungal and oomycete pathogens, also converge on inhibiting apoplastic proteases. The bacterial pathogen *P. syringae* pv. *tomato* DC3000, for example, produces the chagasin-like protease inhibitor Cip1 which was shown to efficiently inhibit the tomato PLCPs C14 and Pip1, and weakly inhibit Rcr3 (Shindo *et al.*, 2016). Like *Pi*EPIC1 an *Pi*EPIC2B, the inhibition of Rcr3 by Cip1 does not trigger *Cf-2*-dependent immunity (Shindo *et al.*, 2016). *Cip1* is broadly conserved in *Pseudomonas* species

and is required for full virulence to tomato (Shindo *et al.*, 2016), indicating that this protease inhibitor is potentially required for virulence in diverse host-pathogen interactions. Finally, the nematode (*Globodera rostochiensis*) effector GrVAP1 was also shown to interact with and inhibit Rcr3 (Lozano-Torres *et al.*, 2012). Surprisingly, this interaction does trigger a *Cf-2*-dpeendent immune response which provides resistance to this nematode (Lozano-Torres *et al.*, 2012).

The fact that Rcr3 can interact with unrelated and structurally divergent inhibitors triggering *Cf-2*-dependent immune responses opens a venue for the rational engineering of this mechanism for an expanded recognition spectrum. However, prior to such rational engineering it is crucial to understand the evolution of the constituent components, as well as advance the understanding of the molecular mechanism.

The aim of this thesis is to characterize the evolution and mechanism of a pathogen recognition mechanism

Chapter 2 discusses the general molecular mechanisms by which resistance genes function. To do this I did a meta-analysis of all cloned *R* genes, thereby identifying nine molecular mechanisms by which they function, substantiated by many examples. This chapter was published in The Plant Cell under the title *" Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function*" (Kourelis and van der Hoorn, 2018). This chapter contains two important advances: i) an unprecedented survey of cloned *R* genes and the molecular mechanism by which they function as well as the pathogens they confer immunity to. This highlights that only for a few *R* genes we know the molecular mechanisms as well as their evolutionary trajectories. A more thorough understanding of mechanisms as well as evolution is required to rationally engineer these genes as well as deploy them in crop species. And, ii) by finding general mechanisms by which these *R* genes work we provide a framework for thinking, which aids in the experimental analysis.

Chapter 3 is focussed on the evolution of the Cf-2/Rcr3 perception mechanism of Avr2. By establishing a transient screening assay, I show that only Rcr3 homologs from species closely related to tomato can trigger an Avr2-dependent HR in the presence of *Cf-2*. This is despite the interaction of these Rcr3 homologs with Avr2, which we show is determined by three critical residues. While *Rcr3* is a highly conserved gene in the Solanaceae, we show that *Cf-2* has evolved very recently in a tomato-specific gene family. This highlights how the evolution of a receptor can confer novel properties to an ancient gene. This dataset is written up for publication under the title "*Evolution of a guarded decoy protease and its receptor in Solanaceae*".

Chapter 4 acts as a bridging chapter. It is a commentary which was published in Trends in Plant Science under the title " *Decoy engineering: the next step in resistance breeding*" (Kourelis *et al.*, 2016). Here we discuss a Science article which describes the rational engineering of the receptor-like cytoplasmic kinase PBS1 to expand the recognition spectrum of the Arabidopsis receptor RPS5 (Kim *et al.*, 2016a). This rational engineering would not have been possible without a thorough understanding of the molecular mechanism by which RPS5/PBS1 recognize the *P. syringae* effector AvrPphB.

Chapter 5 discusses our attempts towards engineering of the Cf-2/Rcr3 mechanism. To engineer this mechanism, we first need to understand how Cf-2 and Rcr3 function to trigger an Avr2-dependent HR. To do this we first engineered the Rcr3 homologs from *N. benthamiana* and eggplant (*Solanum menlongena*) which cannot trigger *Cf-2*-dependent recognition of Avr2 to be able to participate in this recognition. This highlighted several important residues involved in this interaction. Next, we engineered Pip1, the putative operative target of Avr2, to participate in *Cf-2*-dependent recognition of Avr2. This shows that it is possible to substitute Rcr3 for other PLCPs with minimal mutations. Rcr3 and Pip1 have distinct substrate specificities, and an open question is whether we can uncouple substrate specificity from Avr2

recognition. If this is the case it is possible that engineering other PLCPs to interact with *Cf-2* could confer recognition of unrelated proteins. Finally, several attempts at co-immunoprecipitation of Cf-2, Rcr3, and Avr2 were made to try to establish a direct interaction between Cf-2 and Rcr3. This dataset is in preparation for publication.

Finally, **Chapter 6** describes the apoplastic proteome of *Nicotiana benthamiana*, an important model organism in plant pathology. This study involved the re-annotation of the *N. benthamiana* genome. This manuscript is in preparation for re-submission at BMC Biology and is currently online on BioRxiv under the title "*Re-annotated Nicotiana benthamiana gene models for enhanced proteomics and reverse genetics*" (Kourelis *et al.*, 2018).

Chapter 7 discusses the results and perspectives of the five experimental chapters.

Chapter 2

Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function

Kourelis, J. and van der Hoorn, R.A.L. (2018) Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *Plant Cell*, **30**, 285-299

Authors

Jiorgos Kourelis and Renier A.L. van der Hoorn[1]

Affiliations

The Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, OX1 3RB Oxford, United Kingdom

[1]Address correspondence to renier.vanderhoorn@plants.ox.ac.uk.

Author contributions

JK conceived the research with feedback from RALvdH. JK performed the literature searches and developed the storyline with feedback from RALvdH. JK generated the graphs in figure 1 which were subsequently modified by RALvdH with feedback from JK. RALvdH generated figure 2 with feedback from JK. JK wrote the manuscript with feedback from RALvdH.

Justification

In this introductory chapter it will become apparent that resistance to pathogens is often governed by often polymorphic resistance (*R*) genes. Here I summarize the molecular mechanisms underlying the functions of these *R* genes, based on comprehensive meta-analysis of the literature. My subsequent research in chapter 3, 4, and 5 focuses on one such recognition mechanism: indirect recognition of pathogens in the apoplast. To study this mechanism we investigate the recognition of the *Cladosporium fulvum*-derived Avr2 protein by the tomato receptor-like protein Cf-2 which functions together with the tomato papain-like cysteine protease Rcr3.

Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function

Jiorgos Kourelis and Renier A.L. van der Hoorn[1]

Affiliations

The Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, OX1 3RB Oxford, United Kingdom

[1]Address correspondence to renier.vanderhoorn@plants.ox.ac.uk.

Abstract: Plants have many, highly variable resistance (*R*) gene loci, which provide resistance to a variety of pathogens. The first *R* gene to be cloned, maize (*Zea mays*) *Hm1*, was published over 25 years ago, and since then, many different *R* genes have been identified and isolated. The encoded proteins have provided clues to the diverse molecular mechanisms underlying immunity. Here, we present a meta-analysis of 314 cloned *R* genes. The majority of *R* genes encode cell surface or intracellular receptors, and we distinguish nine molecular mechanisms by which R proteins can elevate or trigger disease resistance: direct (1) or indirect (2) perception of pathogen-derived molecules on the cell surface by receptor-like proteins and receptor-like kinases; direct (3) or indirect (4) intracellular detection of pathogen-derived molecules by nucleotide binding, leucine-rich repeat receptors, or detection through integrated domains (5); perception of transcription activator-like effectors through activation of executor genes (6); and active (7), passive (8), or host reprogramming-mediated (9) loss of susceptibility. Although the molecular mechanisms underlying the functions of *R* genes are only understood for a small proportion of known *R*

genes, a clearer understanding of mechanisms is emerging and will be crucial for rational engineering and deployment of novel *R* genes.

INTRODUCTION

Plants can trigger an effective immune response to a wide variety of fungal, bacterial, and viral pathogens. The robustness of the plant immune system is remarkable, given that plants are sessile and unremittingly exposed to potential pathogens while their immunity does not rely on mobile immune cells and adaptive somatic variation, both of which are so effective in vertebrates.

In plants, resistance (*R*) genes play a key role in their remarkable immune responses. *R* genes are usually dominant (but sometimes recessive) genes that provide full or partial resistance to one or more pathogens. We include receptors of pathogen-associated molecular patterns (PAMPs) as *R* genes because they provide partial and sometimes even full resistance (Lacombe *et al.*, 2010). *R* genes exist in natural plant populations and have been used by humankind since early crop domestication. Selection during domestication favoured dominant *R* genes providing full resistance, but recessive *R* genes and *R* genes that provide partial resistance may provide more durable resistance. Most identified *R* genes are polymorphic in plant populations, which led to their initial characterization and use in plant breeding programs. However, individual plants have up to a few hundred *R* gene analogues that make no identified contribution to resistance. Many of these *R* gene analogues are also fixed in plant species and are thought to contribute to non-host resistance (Schulze-Lefert and Panstruga, 2011).

The cloning of the first *R* gene was published in 1992 and the number of cloned *R* genes has steadily increased over the subsequent 25 years (**Figure 1A**). The first cloned *R* gene, *Hm1* from maize (*Zea mays*), encodes an enzyme that detoxifies *Helminthosporium carbonum* (HC) toxin from the fungal pathogen *Cochliobolus carbonum* (Johal and Briggs, 1992). The cloning

of *Hm1* was followed by the cloning of *Pto* from tomato (*Solanum lycopersicum*) in 1993 (Martin *et al.*, 1993), and then, in 1994, *Cf-9* from tomato (Jones *et al.*, 1994), *N* from tobacco (*Nicotiana tabacum*; (Whitham *et al.*, 1994)), and *RPS2* from *Arabidopsis thaliana* (Bent *et al.*, 1994; Mindrinos *et al.*, 1994). Since then, hundreds of examples of cloned *R* genes have been published.

Now, after 25 years of *R* gene cloning, we look back and determine mechanistic trends in the function of *R* gene products. Through a comprehensive review, we identified 314 cloned functional *R* genes (**Supplemental Table 1**). A mechanism has been suggested for 128 of the 314 identified R gene products (41%; **Table 1**); among these, we distinguish nine distinct mechanisms (**Table 1**, **Figure 1A**), explained further in this review, and supported by examples.

Many *R* genes confer recognition of pathogen-derived effectors and initiate effector-triggered immunity (ETI), which often involves the hypersensitive response (HR), a type of programmed cell death (Jones and Dangl, 2006). ETI is considered distinct from the other layer of induced immunity, PAMP-triggered immunity (PTI) (Jones and Dangl, 2006). In contrast to ETI, PTI involves recognition of conserved PAMPs and does not involve HR. In this review, we do not distinguish between PTI and ETI, as the PTI and ETI responses are very similar and the division has become hazy with the description of conserved effectors and polymorphic PAMPs (Thomma *et al.*, 2011).We do, however, distinguish between extracellular and intracellular pathogen recognition mechanisms.

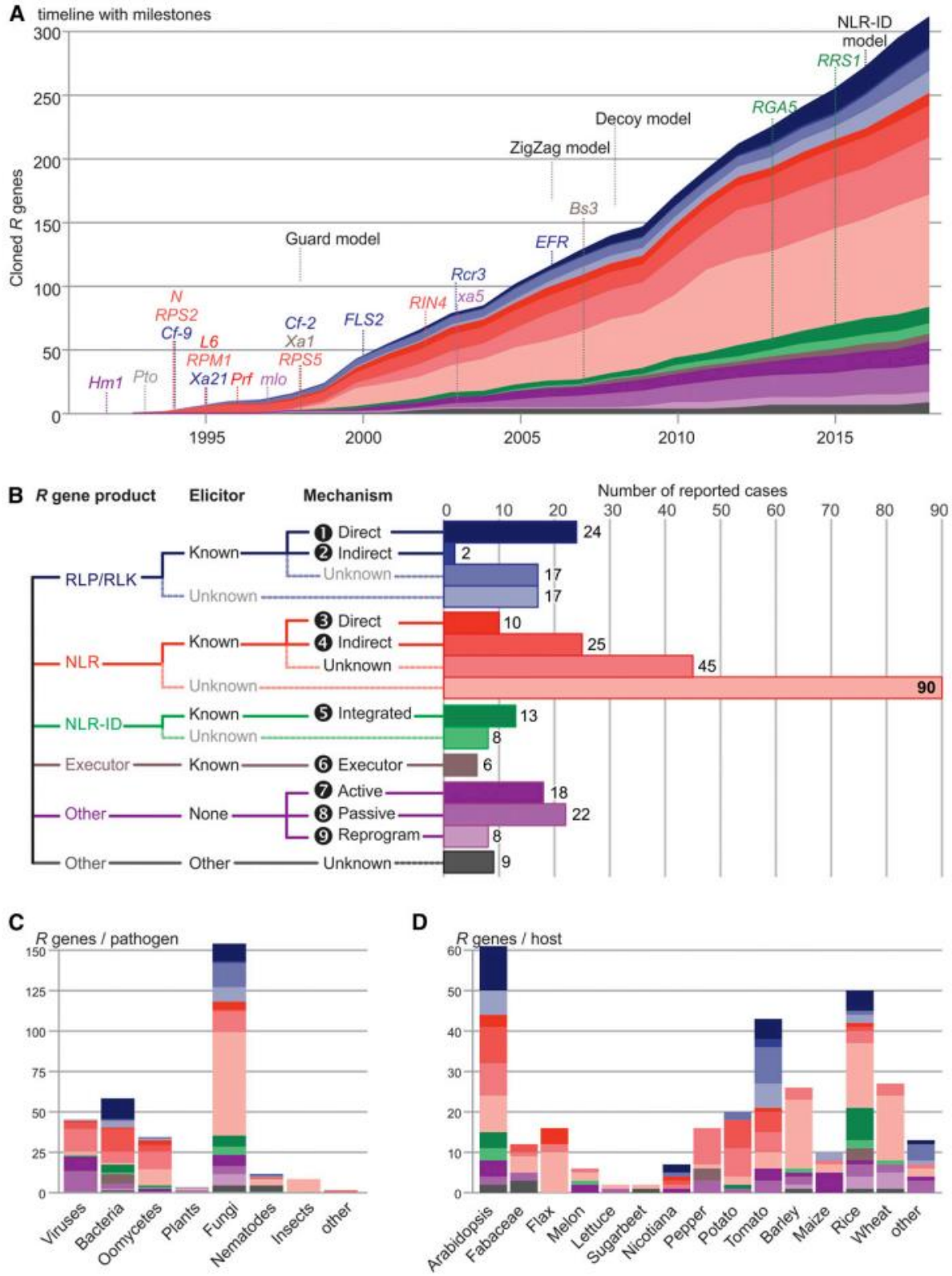


Figure 1. Analysis of Molecular Mechanisms Underpinning R Genes. A near-comprehensive literature search resulted in a database of *R* genes conferring immunity to a wide range of plant pathogens (Supplemental Data Set 1). This list was filtered to remove duplicates: *R* genes that code for receptors that perceive pathogen components were counted only once if this component is conserved between pathogens. By contrast, if an *R* gene codes for a receptor that perceives structurally or sequence unrelated pathogen-derived components,

Figure 1. (continued). these were counted multiple times. Furthermore, alleles of *R* genes with slightly different recognition spectra were counted as multiple *R* genes, as were paralogs within species and orthologues in different species which are shown to be involved in immunity. Loss-of-susceptibility *R* genes that can act against several pathogens through an identical mechanism were counted only once. In addition, *R* genes that were shown to induce enhanced disease resistance when taken outside of their native context were removed, as were engineered *R* genes. Finally, some genes that are involved in autoimmunity resemble *R* genes, but these were not included unless these were proven to be involved in immunity against pathogens. **(A)** A timeline summarizing the increased knowledge regarding *R* genes. The identified mechanisms were plotted cumulatively over time, with some of the milestones added. The date of the first cloned *R* gene was taken in case multiple host components are involved or the underlying molecular mechanism was elucidated later. These *R* genes were grouped according to the proposed mechanism of function as we understand it today. **(B)** All identified *R* genes were grouped according to their proposed mechanism. **(C)** The identified *R* genes were grouped by pathogen which they act against and coloured by the molecular mechanism by which they function. **(D)** The identified *R* genes were grouped by host species carrying them and coloured by the molecular mechanism by which they function. Colours are explained in **(B)**.

Overview and occurrence of the nine mechanisms

Although many cloned *R* genes have no proposed mechanism in the literature, they most likely fit within these nine mechanisms (**Figure 1B**). These mechanisms can broadly be divided into two clades: perception (I) and loss of susceptibility (II). Perception-based mechanisms can be further subdivided into three subclades: perception involving receptor-like proteins/kinases (RLPs/RLKs) (Ia); Nod-like receptors (NLRs) (Ib);and Executor genes (Ic).RLPs/RLKs in the apoplast can perceive PAMPs and effectors directly (subclade Ia, mechanism **1**) or indirectly (subclade Ia, mechanism **2**). Likewise, NLRs can perceive effectors directly (subclade Ib, mechanism **3**) or indirectly (subclade Ib, mechanism **4**), or through integrated domains (IDs) (subclade Ib, mechanism **5**). Perception of transcription activator-like (TAL) effectors occurs by triggering expression of executor genes (subclade Ic, mechanism **6**).

We identify three loss-of-susceptibility mechanisms (clade II): active loss of susceptibility by the expression of immune-related components or specific enzymes involved in detoxification/degradation of pathogen and pathogen-associated components (clade II, mechanism **7**); passive loss of susceptibility through loss of interaction with key pathogen

targets (clade II, mechanism **8**); and loss of susceptibility by host-reprogramming leading to reduced pathogen growth (clade II, mechanism **9**).

Of the cloned *R* genes, 61% encode NLRs (191/313), but for most NLRs the perception mechanism is either unclear (45/191) or the perceived component is unknown (98/191) (**Figure 1B**). The other main class of cloned *R* genes (19%, 60/314), encode RLPs/RLKs, which can recognize pathogen-derived components directly (24/60) or indirectly (2/60).Finally, several cloned *R* genes are executor genes (6/314) or lead to loss of susceptibility (48/314) (**Figure 1B**). Interestingly, with the exception of mechanism 6 (executor genes), the mechanisms are not specific for immunity against certain types of pathogens (**Figure 1C**) or only used by particular plant species or taxonomic groups (**Figure 1D**). Instead, these mechanisms are deployed against different kinds of pathogens and appear to be universally used in the plant kingdom. However, the distribution of these mechanisms may not reflect the natural frequencies of mechanisms because of a strong research bias toward model systems.

Subclade Ia: Extracellular Perception

Plants perceive various PAMPs and effectors by *R* genes encoding for cell surface-localized RLPs or RLKs (**Figure 2**). This perception can be direct (mechanism **1**) or indirect (mechanism **2**).

Mechanism 1: Direct Perception at the Cell Surface

Many PAMPs are perceived directly at the cell surface by RLPs and RLKs. The best studied PAMP in plants is bacterial flagellin (Felix *et al.*, 1999). Flagellin fragment flg22 is perceived by the RLK FLAGELLIN-SENSITIVE2 (FLS2) in Arabidopsis (Gómez-Gómez and Boller, 2000). The flg22 peptide binds directly to FLS2 triggering the recruitment of BRI1-ASSOCIATED RECEPTOR KINASE (BAK1), a RLK (Chinchilla *et al.*, 2007). The crystal

Table 1. *R* Genes Classified by Nine Mechanisms

Mechanism	Description	*R* Genes (Plant Species)
1: RLP/RLK, direct	Recognition triggered by direct interaction of a pathogen-derived effector and a cell surface RLK/RLP receptor.	*EFR*, ***FLS2***, *LORE*, *LYK3*, *LYK4*, *LYK5*, *LYM1*/*LYM3*, *LYM2*, *RBGP1*, *RLP23* (Arabidopsis), *VvFLS2* (grapevine), *NbCORE*, *NbFLS2* (N. benthamiana), *CEBiP*, *LYP4*/*LYP6*, *OsFLS2* (rice), *CORE*, *FLS3*, *LeEIX2*, *SlFLS2* (tomato)
2: RLP/RLK, indirect	Recognition triggered either by effector binding to a host component or by effector-mediated modification of a host component, perceived by a cell surface RLK/RLP receptor.	***Cf-2*** (tomato)
RLP/RLK, unknown mechanism		*Ve1* (tomato), *StoVe1* (eggplant), *HLVe1-2A* (hop), *NgVe1* (Nicotiana glutinosa), *LepR3*, *RLM2* (oilseed rape), *StuVe1*, *ELR* (potato), *XA21* (rice), *9DC1*, *9DC2*, *9DC3*, *Cf-4*, *Cf-5*, *Cf-9*, *Hcr9-4E*, *I*, *I-3* (tomato)
3: NLR, direct	Recognition triggered by direct interaction of a pathogen-derived component and an NLR.	***RPP1***-{*EstA*/*Nda*/*ZdrA*} (Arabidopsis), ***L5***/***L6***/***L7***, *M* (flax), *Roq1* (N. benthamiana), *Pi-ta* (rice), *Sw-5b* (tomato)
4: NLR, indirect	Recognition triggered either by effector binding to a host component or by effector-mediated modification of a host component, perceived by an NLR.	*HRT1*, ***RPM1***, ***RPS2***, ***RPS5***, ***SUMM2***, ***ZAR1*** (Arabidopsis), *Gpa2*, *R2*, *R2-like*, *Rpi-abpt*, *Rpi-blb3*, *Rx1*, *Rx2* (potato), *Rpg1-b*, *Rpg1r* (soybean), *N* (tobacco), *Prf* (tomato) Guardees/decoys: *TIP*, *RIN4*, *PBS1*, *CRCK3*, *ZED1*, *ZRK3*, *RKS1*, *PBL2* (Arabidopsis), *RanGAP2*, *BSL1* (soybean), *GmRIN4* (tobacco), *LescPth5*, *Fen*, *Pto* (tomato)
NLR, unknown mechanism		*RBA1*, *RCY1*, *RPP13-Nd-1*, *RPP13-UKID37*, *RPP13-UKID5*, *RPP5*, *RPS6*, *TAO1* (Arabidopsis), *Mla1*, *Mla10*, *Mla13* (barley), *Bs2* (black pepper), *P*, *P2* (flax), *Rxo1* (maize), *Fom-2* (melon), L^{1}, L^{1a},L^{1c}, L^{2}, L^{2b}, L^{3}, L^{4}, *Pvr4*, *Tsw* (pepper), *R3a*, *R3b*, *R8*, *Rpi-blb2*, *Rpi-vnt1.1*, *Rpi-vnt1.2*, *Rpi-vnt1.3* (potato), *Pi9*, *Pib*, *Piz-t* (rice), *3gG2* (soybean), *N'* (tobacco), *Bs4*, *I2*, *Tm-2*, $Tm2^{2}$ (tomato), *Pm2*, *Pm3a*, *Pm3f* (wheat)
5: NLR-ID	Recognition is triggered either by effector binding to a domain or by effector-mediated modification of a domain that is integrated in a host NLR.	*RRS1B*, ***RRS1-R***, *RRS1-S* (Arabidopsis), *R1* (potato), *Pii-2*, ***Pik***-{*1*/*h*/*p1*/*s*}, ***RGA5-A***, *Xa1* (rice)
6: Executor	Recognition triggered by transcriptional activation of the executor gene by a pathogen TAL effector.	***Bs3***, *Bs3-E*, *Bs4C-R* (pepper), *Xa10*, *Xa23*, *Xa7* (rice)
7: Other, active	Loss of susceptibility by directly disarming the pathogen by actively interrupting a key pathogenicity process.	*JAX1*, *RTM1*, *RTM2*, *RTM3* (Arabidopsis), *HvHm1* (barley), ***Hm1***, *Hm2*, *qMdr9.02*, *qRfg1*, *ZmTrxh* (maize), *At1*, *At2* (melon), *IVR* (N. benthamiana), *PaLAR3* (Norway spruce), *STV11-R* (rice), *Tm-1*, *Ty-1*, *Ty-3* (tomato)
8: Other, passive	Loss of susceptibility by mutation in a host component, leading to the inability to manipulate the host.	*rwm1*, l*ov1* (Arabidopsis), *Rym-4*, *Rym-5* (barley), *retr01* (cabbage), *bc-3* (French bean), *Mo-1* (lettuce), *sbm1* (pea), $Pvr2^{1}$, $Pvr2^{2}$, *pvr6* (pepper), *Eva1* (potato), *xa13*, *xa25*, ***xa5*** (rice), *LGS1*, *Pc* (sorghum), *Asc-1*, *Ty-5*, *Pot-1* (tomato), *Tsn1*, *Snn1* (wheat)
9: Reprogramming	Loss of susceptibility by a deregulated host.	***mlo*** (barley), *GH3-2*, *GH3-8*, *pi21* (rice), *Lr34*, *Lr67*, *YrL693*, *Yr36* (wheat)

Several *R* genes have multiple synonyms. This table summarizes only one name per cloned *R* gene. Only the name of the sensor NLR is mentioned in case of NLR-IDs. *R* genes that can perceive multiple sequence-unrelated effectors have been counted multiple imes. Bold indicates key examples discussed in this work.

structure of FLS2 in complex with BAK1 and flg22 has been elucidated, revealing that flg22 acts as a "molecular glue" inducing complex formation of FLS2 with BAK1 (Sun *et al.*, 2013). FLS2 orthologues from different plant species perceive different parts of bacterial flagellin. For instance, tomato SlFLS2 recognizes a 15-amino acid epitope that overlaps with flg22 (Robatzek *et al.*, 2007), whereas different flagellin epitopes (flgII-28 and CD2-1) are

recognized by the RLK FLS3 in tomato and an uncharacterized receptor in rice (*Oryza sativa*), respectively (Hind *et al.*, 2016; Katsuragi *et al.*, 2015). Furthermore, FLS2 receptors from different plant species display different affinities for the conserved part of flagellin from different pathogens, possibly reflecting coevolution with specific bacterial pathogens (Trdá *et al.*, 2014).

Many other microbial PAMPs are recognized directly similar to flagellin (reviewed in Zipfel, 2014). These include EF-Tu, chitin, lipopolysaccharide, peptidoglycan, and other components of the bacterial cell wall (Zipfel, 2014), as well as bacterial RNA (Lee *et al.*, 2016), possibly by direct interaction with RLP/RLKs. Although often not considered as being *R* genes, PAMP-recognizing RLKs provide partial resistance and can even confer full resistance when transferred to other plant species (Lacombe *et al.*, 2010).

Perception of extracellular effector proteins may also occur via direct interactions with RLPs or RLKs. For example, nlp20, a fragment of NECROSIS AND ETHYLENE-INDUCING PEPTIDE1 (NEP1), has been co-immunoprecipitated with RLP23 (Albert *et al.*, 2015). NEP-LIKE PROTEINs (NLPs) are secreted by bacterial, oomycete, and fungal plant pathogens (Böhm *et al.*, 2014; Oome *et al.*, 2014). Many NLPs function in the necrotrophic phase of these pathogens and have a cytotoxic role. Non-cytotoxic NLPs, on the other hand, are expressed early during infection and serve an unknown role during the initial stages of host colonization (Oome *et al.*, 2014). Both cytotoxic (Böhm *et al.*, 2014) and non-cytotoxic NLPs (Oome *et al.*, 2014) trigger immune responses via RLP23 through the conserved nlp20 fragment (Albert *et al.*, 2015). RLP23 requires the RLKs SUPPRESSOR OF BIR1 1 and BAK1 for signaling, triggering a weak immune response (Albert *et al.*, 2015).

A direct interaction has been demonstrated for the perception of the fungal ethylene-inducing xylanase (EIX) effector by the tomato RLP LeEIX2 (Ron and Avni, 2004; Rotblat *et al.*, 2002).

Expression of LeEIX2 in mammalian cells enables binding of EIX (Ron and Avni, 2004). Furthermore, several fungal endopolygalacturonases are perceived by the Arabidopsis RLP RBPG1 (Zhang *et al.*, 2014), and these enzymes can be co-immunoprecipitated with RBPG1 (Zhang *et al.*, 2014). In addition, perception of EIX and fungal endopolygalacturonases by LeEIX2 and RBPG1, respectively, is independent of their enzymatic functions (Rotblat *et al.*, 2002; Zhang *et al.*, 2014). A direct interaction is thought to also occur for the RaxX effector from Xanthomonas oryzae, which is perceived by the RLK encoded by the rice *R* gene Xa21 (Pruitt *et al.*, 2017). Likewise, the Verticillium-derived effector AVE1 is perceived by the tomato R gene Ve1, which encodes an RLP (de Jonge *et al.*, 2012). However, no direct interaction has been shown for these effectors and their cognate Xa21 and Ve1 receptors.

Taken together, these examples indicate that direct perception of PAMPs and effectors at the cell surface is a common strategy underlying the partial and full disease resistance provided by *R* genes.

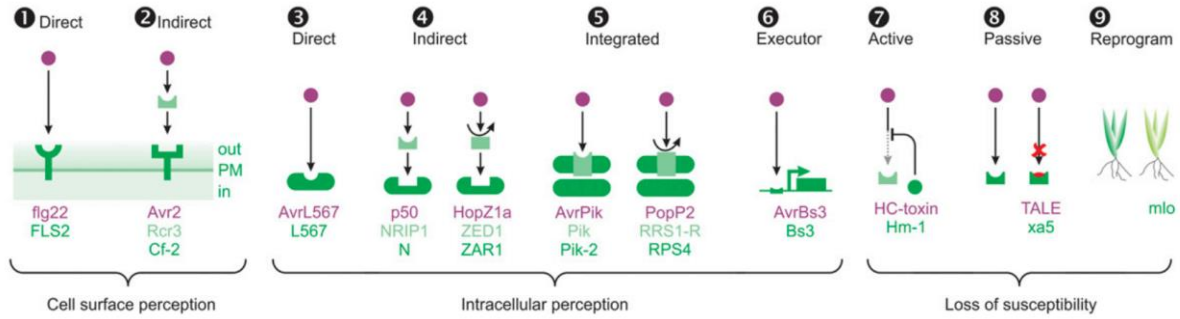


Figure 2. Nine Molecular Mechanisms Underpinning R Gene Functions. Illustration of direct (1) and indirect (2) recognition at the cell surface; four different intracellular perception mechanisms (3–6); and three loss-of-susceptibility mechanisms (7–9). PAMPs and effectors are coloured in purple, indirect receptors in light green, and direct receptors in dark green.

Mechanism 2: Indirect Extracellular Effector Perception

Pathogen perception at the cell surface can also occur indirectly, by recognition of modified host factors. The key example of this mechanism is the recognition of the fungal pathogen

Cladosporium fulvum (syn. *Passalora fulva*) by the tomato *R* gene product Cf-2, a RLP which has been introgressed into cultivated tomato from *Solanum pimpinellifolium* (Dixon *et al.*, 1996). *Cf-2* confers recognition of the fungal effector Avr2 from C. fulvum (Luderer *et al.*, 2002), as well as the unrelated nematode effector GrVap1 from Globodera rostochiensis (Lozano-Torres *et al.*, 2012). Avr2/GrVap1 perception requires *Rcr3*, which encodes a secreted papain-like Cys protease (Dixon *et al.*, 2000). Both Avr2 and GrVAP1 function as protease inhibitors and directly interact with Rcr3 (Lozano-Torres et al., 2012; Rooney et al., 2005). Avr2 and GrVap1 act as virulence factors (van Esse *et al.*, 2008; Lozano-Torres *et al.*, 2014), likely by inhibition of various host proteases (Kaschani *et al.*, 2010; Lozano-Torres *et al.*, 2012; 2014; Shabab *et al.*, 2008), including the Rcr3 paralog Pip1 (Shabab *et al.*, 2008). These data suggest that Rcr3 might act as a decoy for the perception of Avr2 and GrVap1 (van der Hoorn and Kamoun, 2008). However, tomato lines lacking a functional *Rcr3* allele are more susceptible to *Phytophthora infestans*, but do not display an altered susceptibility toward *C. fulvum* or *Pseudomonas syringae* (Ilyas *et al.*, 2015). This indicates a role for Rcr3 beyond effector recognition. A second argument against Rcr3 being a decoy is that close Rcr3 orthologues are present in pepper (*Capsicum annuum*), potato (*Solanum tuberosum*), and tomato (Ilyas *et al.*, 2015), whereas *Cf-2* is probably an evolutionarily much younger gene present only in tomato (Dixon *et al.*, 1996; Seear and Dixon, 2003).

So far, Avr2/GrVap1 perception by Cf-2 and Rcr3 is the only example of indirect perception of effectors in the apoplast. However, perception of other effectors in the apoplast might also be indirect. For example, a direct interaction of tomato RLP Cf-9 with the *C. fulvum* effector Avr9 could not be detected (Luderer *et al.*, 2001), and Avr9 perception seems to involve an additional Solanaceae-specific factor known as the high-affinity binding site for Avr9 (Kooman-Gersmann *et al.*, 1996; 1998). Likewise, perception of the *Fusarium oxysporum*

Avr3 effector by the tomato RLK I-3 may require an additional host component present in tomato, but absent in tobacco and *Nicotiana benthamiana* (Catanzariti *et al.*, 2015).

Subclade Ib: Intracellular Perception

In addition to extracellular perception of pathogens, plants are able to recognize effectors inside the cell (**Figure 2**). Most cloned *R* genes encode NLRs, which can perceive effectors either directly (mechanism **3**) or indirectly (mechanism **4**). However, some *R* genes encode NLRs containing additional IDs (NLR-IDs) required for effector recognition (mechanism **5**) (Kroj *et al.*, 2016; Sarris *et al.*, 2016). Finally, some effectors activate executor genes (mechanism **6**).

Mechanism 3: Direct Intracellular Effector Recognition—Trigger-Happy NLRs

Direct perception of effectors is not restricted to the cell surface, as several effectors have been convincingly demonstrated to directly interact with NLRs to trigger immune responses. ARABIDOPSIS THALIANA RECOGNIZED1 (ATR1), for example, is an effector from the oomycete pathogen *Hyaloperonospora arabidopsidis*, and ATR1 directly interacts with the Arabidopsis NLR RECOGNITION OF PERONOSPORA PARASITICA1 (RPP1), leading to its recognition (Krasileva *et al.*, 2010). Various *RPP1* alleles have distinct recognition specificities for different *ATR1* alleles, and this specificity is mediated by the LRRs of RPP1 (Steinbrenner *et al.*, 2015). The AvrL567 effector from the fungal pathogen *Melampsora lini* is another effector that directly interacts with its cognate receptor, the flax L5, L6, and L7 NLRs, which are encoded by allelic genes (Dodds *et al.*, 2004; 2006). Different variants of AvrL567 are perceived differentially by these receptors (Dodds *et al.*, 2006; Wang *et al.*, 2007), and the specificity for these AvrL567 variants resides in the LRR domain of the NLR (Ravensdale *et al.*, 2012; Wang *et al.*, 2007).

Interestingly, effector perception is also affected by cooperative polymorphisms in other domains of the NLR besides the LRR (Ravensdale *et al.*, 2012). This suggests a mechanism of

perception where effector binding competes with intramolecular interactions (Ravensdale *et al.*, 2012). The L5, L6, and L7 receptors exist in an equilibrium between "on" and "off" states, and effector binding stabilizes the "on" form, thereby initiating immune signaling (Bernoux *et al.*, 2016). This equilibrium model may be more widely applicable to NLR signaling. Indeed, the NLR Sw-5b directly recognizes a small 21-amino acid PAMP-like peptide region from the viral movement protein NSm (NSm^{21}), and it has been argued that binding of NSm^{21} destabilizes the interaction between the LRR and NB-ARC domain triggering recognition (Zhu *et al.*, 2017).

In the case of RPP1 and L5/L6/L7, allelic NLR-encoding genes perceive allelic effector variants. However, it is not always the case that homologous and/or allelic NLRs perceive homologous effector proteins. For example, rice and barley (*Hordeum vulgare*) have homologous NLRs residing in the same *R* gene cluster; these NLRs perceive sequence-unrelated effectors from *Magnaporthe oryzae* and *Blumeria graminis,* respectively (Lu *et al.*, 2016; Wu *et al.*, 2015). In addition, potato orthologues of two different tomato NLR-encoding genes, tomato *Sw-5b*, which perceives a viral effector, or tomato *I2*, which perceives a fungal effector, can perceive oomycete effectors (Giannakopoulou *et al.*, 2015; Vossen *et al.*, 2016). The exact molecular mechanism by which sequence-unrelated effectors from different pathogens can be perceived directly by highly similar NLRs is unknown. In many cases, the specificity-determining region in the NLR is in the C-terminal LRR region (Shen *et al.*, 2003), while in other cases mutations in regions other than the NLR are sufficient to alter specificity (Giannakopoulou *et al.*, 2015). Therefore, it seems that small changes in several parts of the NLR can sensitize the NLR to destabilization by sequence-unrelated effectors, triggering recognition. In addition, evolving a different specificity to sequence-unrelated effectors may be facilitated by the fact that these effectors might adopt a similar fold. Indeed, several sequence-unrelated effectors from *M. oryzae* have been shown to adopt a similar fold (de

Guillen *et al.*, 2015; Maqbool *et al.*, 2015). A structural understanding of effector binding by these NLRs would provide clues to the exact molecular mechanism of direct perception, but these experiments are notoriously challenging.

Mechanism 4: Indirect Intracellular Recognition—Decoys and Guardees

Many effectors are perceived indirectly by NLRs, through their interaction with—or enzymatic modification of—additional host components. These additional host proteins have also been called guardees or decoys to suggest that they are effector targets or mimics of effector targets, respectively.

An example of indirect interaction is the recognition of the 50-kD helicase (p50) domain of Tobacco mosaic virus by the tobacco NLR N, which requires NRIP1, a chloroplast-localized rhodanese sulfurtransferase (Caplan *et al.*, 2008). NRIP1 is recruited to the cytoplasm by the p50 effector, and NRIP1 only interacts with N in the presence of the p50 effector. It is thought that N activation requires a pre-recognition complex of NRIP1 and the p50 effector (Caplan *et al.*, 2008).

An example of indirect detection via enzymatic host modification is the perception of the *P. syringae* Type-III effectors HopZ1a (Lewis *et al.*, 2010) and HopF2a (Seto *et al.*, 2017), and the *Xanthomonas campestris* Type-III effector AvrAC (syn. XopAc) (G., Wang *et al.*, 2015; Xu *et al.*, 2008) by the functionally conserved NLR ZAR1 (Baudin *et al.*, 2017; Lewis *et al.*, 2010; Peele *et al.*, 2014). HopZ1a is perceived through its acetyltransferase activity on the PM-localized pseudokinase HOPZ-ETI-DEFICIENT1 (ZED1), a receptor-like cytoplasmic kinase (RLCK) XII family member (Lewis *et al.*, 2013). AvrAC is perceived through its activity on the RLCK VII family member PBL2 (G., Wang *et al.*, 2015). When PBL2 is uridylylated by AvrAC, it interacts with the ZED1-like RKS1, which in turn forms a stable complex with HOPZ-ACTIVATED RESISTANCE1 (ZAR1), leading to a weak HR (G., Wang *et al.*, 2015).

HopF2a is an ADP-ribosyltransferase and its recognition requires ZAR1 and the ZED1-like ZRK3, which may act as an adaptor for effector recognition similar to RKS1 (Seto *et al.*, 2017). ZED1 and PBL2 are thought to act as decoys for the perception of HopZ1a or AvrAC, respectively. Indeed, *ZED1* does not contribute to immunity or susceptibility to *P. syringae* in the absence of ZAR1, indicating that ZED1 acts as a decoy (Lewis *et al.*, 2013). By contrast, PBL2 is not a strict decoy, as it appears to be required for perception of certain PAMPs, including flg22 and elf18 (Zhang *et al.*, 2010). It is yet undetermined whether *PBL2* contributes to resistance or susceptibility to *X. campestris* in the absence of *ZAR1*. The exact molecular mechanism by which ZAR1 perceives modulation of various RLCKs is unknown. ZAR1 is a classic example of how one *R* gene product can perceive the action of multiple effectors from multiple pathogens through interactions with several host components, all of which may be able to recruit additional components.

Another functionally conserved NLR that indirectly perceives multiple effectors is tomato Prf ((Bombarely *et al.*, 2012; Salmeron *et al.*, 1996); reviewed in Ntoukakis *et al.*, 2014). Prf interacts with several RLCK VII family members, including Pto and Fen (Salmeron *et al.*, 1996). Interaction with Pto leads to perception of the *P. syringae* Type-III effectors AvrPto and AvrPtoB, which bind Pto (Kim *et al.*, 2002). AvrPto acts as a kinase inhibitor (Xing *et al.*, 2007), while AvrPtoB acts as a weak kinase inhibitor and has an additional E3 ligase domain to target its interactors for destruction by the host proteasome (Cheng *et al.*, 2011; Rosebrock *et al.*, 2007). Prf also interacts with the Pto-homolog Fen, leading to immune signaling in response to the insecticide Fenthion (Martin *et al.*, 1994). It is thought that Prf oligomerizes with several molecules of Pto and/or Pto homologs and that effector binding to one of these kinases in this complex would induce a series of transphosphorylation events resulting in an immune response (Ntoukakis *et al.*, 2014). The oligomerization with different Pto homologs

within one receptor complex could expand the recognition spectrum of Prf (Ntoukakis *et al.*, 2014).

The NLRs ZAR1 and Prf each interact with multiple host proteins; by contrast, RPM1 INTERACTING PROTEIN4 (RIN4) is a host protein that interacts with multiple NLRs. RIN4 forms complexes with the NLRs RPS2 and RPM1 (Axtell and Staskawicz, 2003; Mackey *et al.*, 2002). The *P. syringae* Type-III effectors AvrB and AvrRpm1 induce phosphorylation of RIN4 by recruiting the Arabidopsis RIPK protein kinase (J., Liu *et al.*, 2011; Mackey *et al.*, 2002). The phosphorylation of RIN4 by RIPK at a conserved threonine residue (T166) reduces RIN4 interaction with the prolyl-peptidyl isomerase ROTAMASE CYP1 (ROC1) leading to an altered conformation of RIN4, which triggers RPM1-mediated immune signaling (Li *et al.*, 2014). In the absence of RPM1, the phosphorylation of RIN4 enhances its role as a negative regulator of immunity (Lee *et al.*, 2015; Liu *et al.*, 2009; Wilton *et al.*, 2010), indicating RIN4 is not a decoy. Cleavage of RIN4 by the *P. syringae* Type-III effector AvrRpt2 inhibits RPM1-mediated immune signaling (Kim *et al.*, 2005). This cleavage requires AvrRpt2 maturation by eukaryotic prolyl-peptidyl isomerases such as ROC1 (Coaker *et al.*, 2005), which is required for RPM1-mediated immune signaling. In response, the NLR RPS2 perceives RIN4 cleavage, triggering immune signaling (Axtell and Staskawicz, 2003). RIN4 binding to RPS2 keeps RPS2 in an "off" state, and cleavage releases the inhibition of RPS2 triggering immune signaling (Day *et al.*, 2005). In addition, RPS2 weakly perceives AvrRpm1 (Kim *et al.*, 2009). In the absence of *RPM1*, AvrRpt2-dependent cleavage of RIN4 does not contribute to virulence (Lim and Kunkel, 2004), indicating that in the absence of AvrB or AvrRpm1, RIN4 acts as a decoy for AvrRpt2. Soybean (*Glycine max*) has independently converged on a similar perception mechanism, where the soybean NLRs RPG1-b and RPG1r perceive the manipulation of the soybean RIN4 ortholog by AvrB and AvrRpm1, respectively (Ashfield *et al.*, 2004; Kessens *et al.*, 2014; Selote and Kachroo, 2010).

Other NLRs also perceive effectors indirectly using additional host proteins. The NLR RPS5, for example, perceives the *P. syringae* Type-III effector AvrPphB (Warren *et al.*, 1998), a family C58 cysteine protease (Zhu *et al.*, 2004). AvrPphB contributes to virulence by cleaving several RLCKs (Zhang *et al.*, 2010), including RIPK, thereby inhibiting RPM1-mediated recognition of AvrB (Russell *et al.*, 2015). The perception of AvrPphB by RPS5 requires *PBS1*, which encodes a RLCK VII family member (Ade *et al.*, 2007; Shao *et al.*, 2003a; Swiderski and Innes, 2001). PBS1 binds to RPS5 to form a pre-recognition complex (DeYoung *et al.*, 2012), and subsequent cleavage of PBS1 by AvrPphB triggers a subtle conformational change that allows an exposed loop in PBS1 to trigger activation of RPS5 (DeYoung *et al.*, 2012; Dong Qi *et al.*, 2014). It remains to be demonstrated whether PBS1 cleavage by AvrPphB contributes to the virulence of *P. syringae* in the absence of *RPS5* and therefore whether PBS1 is a decoy. However, *PBS1* may not be a decoy, as PBS1 contributes to PAMP-triggered immune responses (Zhang *et al.*, 2010) and is under strong purifying selection, even though not all Arabidopsis accessions carry *RPS5* (Caldwell and Michelmore, 2009), indicating a physiological function for PBS1 in the absence of *RPS5*.

Finally, sometimes NLRs monitor cellular homeostasis. The Arabidopsis NLR SUMM2, for example, perceives the phosphorylation status of the calmodulin binding RLCK CRCK3 (Zhang *et al.*, 2017). CRCK3 is phosphorylated by a MAP kinase signaling cascade involving the MAP kinases MEKK1, MKK1, and MKK2, as well as MPK4 (Zhang *et al.*, 2017). The CRCK3 phosphorylation status in the absence of MPK4, MKK1/MKK2, or MEKK1 kinase activity triggers an immune response that is dependent on SUMM2 (Zhang *et al.*, 2012). The *P. syringae* Type-III effector HopAI1 inactivates MPK4, thereby triggering an immune response that is dependent on SUPPRESSOR OF MKK1 MKK2 2 (SUMM2; (Zhang *et al.*, 2012)). SUMM2 is therefore an example of an NLR monitoring cellular homeostasis. Perception of an alteration of homeostasis, rather than the actual effector target itself, may

allow for the perception of many effectors whose effects ultimately converge on the same host components. CRCK3 has no known function beyond its signaling role through SUMM2, suggesting it might act as a decoy (Zhang *et al.*, 2012).

In conclusion, NLRs often trigger immune signaling by monitoring either (1) the interaction of effectors with host proteins, (2) the enzymatic modification of host proteins, or (3) cellular homeostasis. More bacterial effectors have been shown to be perceived indirectly by NLRs than effectors of other classes of pathogens (**Figure 1C**). This likely reflects a better understanding of bacterial effectors as well as the fact that bacteria are easier to manipulate, rather than a biological difference in how these different types of pathogens are perceived by plants. Although there are several clear candidates acting as guardees (RIN4 and PBL2) or decoys (ZED1 and Pto), it is unclear in other cases (PBS1 and CRCK3). Furthermore, several of these proposed decoys seem more conserved than the corresponding NLR.

Mechanism 5: NLR-IDs—Sensors Embedded within Receptors

Some NLRs carry integrated domains (NLR-IDs) (Kroj *et al.*, 2016; Sarris *et al.*, 2016) thought to be required for effector recognition. As these additional domains seem to have integrated repeatedly and recently, it is not unlikely they still possess their original activity in the absence of the recognized effector, unlike a decoy (Sarris *et al.*, 2016). Furthermore, it remains to be demonstrated on a case-by-case basis whether the ID is required for effector perception. Three NLR-IDs have been studied in detail: Arabidopsis RRS1 and rice RGA5 and Pik. Remarkably, these NLR-IDs are co-expressed from the same promoter in opposing orientation with a NLR lacking the ID, and heterodimerization of these paired, co-expressed NLR proteins is required for signal transduction.

Arabidopsis RRS1 is a NLR containing an integrated C-terminal WRKY transcription factor domain and cooperates with the NLR RPS4 to trigger effector recognition. The WRKY domain

of RRS1 interacts with the *P. syringae* Type-III effector AvrRps4, triggering RPS4-dependent immune responses (Sarris *et al.*, 2015). PopP2 is a *Ralstonia solanacearum* Type-III effector of the YopJ family that acts as an acetyltransferase, acetylating key lysines in the RRS1 WRKY domain. In the Col-0 allele of *RRS1* (*RRS1-S*), this acetylation blocks AvrRps4 recognition (Sarris *et al.*, 2015), while the Nd-1 and Ws-2 alleles of *RRS1* (*RRS1-R*) signal in response to this acetylation, in addition to recognizing AvrRps4 (Le Roux *et al.*, 2015; Sarris *et al.*, 2015). Furthermore, the *RRS1-R* allele recognizes a yet unidentified effector from the fungus *Colletotrichum higginsianum* (Sarris *et al.*, 2015), and the *RPS4-RRS1* locus also underlies a quantitative trait locus for resistance against *X. campestris* (Debieu *et al.*, 2016). Therefore, RRS1 can recognize effectors from various bacterial and fungal pathogens. Furthermore, RRS1-R can perceive the direct interaction of AvrRps4 with its integrated WRKY domain, as well as perceive the enzymatic modification of this domain by PopP2. The biological significance of the RRS1 WRKY domain binding to W-box DNA sequences (Sarris *et al.*, 2015) outside of the context of perceiving pathogen effectors is not known. Remarkably, AvrRps4 is also perceived in a similar way in Arabidopsis by a homologous pair of NLRs, designated RRS1B (containing a WRKY ID), and RPS4B (Saucet *et al.*, 2015).

The rice RGA5 and Pik NLRs both contain an integrated RATX1 or HMA domain, which is related to the *Saccharomyces cerevisiae* copper binding protein ATX1 (Cesari *et al.*, 2013). The HMA domain is located at the C terminus of RGA5 and between the CC and NB-ARC domain in Pik, representing independent integration events. RGA5 interacts with the *M. oryzae* effectors Avr-Pia and Avr-CO39 using its HMA-domain, and this triggers an RGA4-dependent immune response (Cesari *et al.*, 2013). Alleles of the *M. oryzae* effector Avr-Pik bind to the HMA domain of Pik alleles (Maqbool *et al.*, 2015), which together with Pik-2 triggers immune signaling (Ashikawa *et al.*, 2008). Substitutions in the HMA-domain of different Pik alleles

revealed a correlation between the recognition specificity of these different Pik alleles and the different Avr-Pik variants (Maqbool *et al.*, 2015).

In conclusion, perception by NLR-IDs can be the result of direct binding of the effector or due to enzymatic activity of the effector on the integrated domain. In addition, studying NLR-IDs can provide valuable insight into the molecular targets of effectors, as these integrated domains likely represent domains targeted by these effectors (Kroj *et al.*, 2016). All studied NLR-IDs require a genetically linked NLR to trigger signaling. However, genetically unpaired NLRs can also cooperate. In Arabidopsis, tobacco, and tomato, unlinked "helper" NLRs have been shown to act downstream of several canonical "sensor" NLRs (Bonardi *et al.*, 2011; Peart *et al.*, 2005; Wu *et al.*, 2017). In addition, helper-NLRs are required for effector perception by RLPs (Fradin *et al.*, 2009; Gabriëls *et al.*, 2007), thereby linking effector perception in the apoplast to intracellular signaling by NLRs.

Mechanism 6: Executor Genes—Promoter Traps for TAL Effectors

Executor genes are *R* genes that are transcriptionally activated by transcription activator-like effectors (TALEs) produced by *Xanthomonas* species and confer immunity to the *Xanthomonas* strains carrying these TALEs. TALEs function by binding to specific DNA sequences, altering host transcription of key susceptibility factors. Executor genes function as a promoter trap for these TALEs, forcing them to promote transcription of genes involved in immunity. Promoters of executor genes therefore act as decoys, mimicking the promoter regions of these susceptibility factors, leading to the initiation of immune responses (van der Hoorn and Kamoun, 2008). So far, six executor genes have been cloned: rice *Xa27* (Gu *et al.*, 2005), *Xa10* (Tian *et al.*, 2014), and *Xa23* (C., Wang *et al.*, 2015), and pepper *Bs3*/*Bs3-E* (Römer *et al.*, 2007), and *Bs4C-R* (Strauß *et al.*, 2012). Identified executor genes either encode for proteins with a catalytic function (*Bs3* and *Bs3-E* encode for a putative flavin monooxygenase; (Römer *et al.*, 2007)) or for proteins with multiple putative transmembrane

domains (*Xa27*, *Xa10*, *Xa23*, and *Bs4C-R*) (Zhang *et al.*, 2015). Importantly, increased understanding of the specificity of DNA binding by TALEs has enabled the development of synthetic executor genes mediating immunity against multiple *Xanthomonas* strains (Hummel *et al.*, 2012; Römer *et al.*, 2009; Zeng *et al.*, 2015). This strategy could be used to engineer resistance against the pathogen *R. solanacearum*, which also expresses TALEs, known as RipTALs (de Lange *et al.*, 2013).

Clade II: Loss of Susceptibility

Not all *R* genes encode proteins involved in mechanisms relying on perception. Several *R* genes rely on loss of susceptibility, sometimes also called susceptibility (*S*) genes (van Schie and Takken, 2014). Loss of susceptibility can occur in three ways (**Figure 2**): active (mechanism **7**), passive (loss of interaction; mechanism **8**), or by host reprogramming (by mutations in components of cellular pathways; mechanism **9**). Many loss-of-susceptibility mechanisms of resistance are very durable, but can carry a cost for the host and therefore can lead to a yield penalty. When active and passive loss-of-susceptibility mechanisms become fixed in populations, they lead to Type-I non-host resistance, typically characterized by the absence of a strong immune response (Mysore and Ryu, 2004).

Mechanism 7: Active Loss of Susceptibility—Blocking Key Pathogen Strategies

R genes underlying active loss-of-susceptibility mechanisms encode host proteins that are directly involved in disarming the pathogen by actively interrupting a key pathogen process. Active loss-of-susceptibility mechanisms are diverse and can act broadly against many pathogens, or be specific to some. These active mechanisms are often constitutively expressed, but some can be upregulated upon immune signaling. In addition, active loss-of-susceptibility mechanism may also lead to the production of PAMPs and danger-associated molecular patterns (DAMPs), which are then perceived and amplify the original response. For example, *Hm1*, the first *R* gene cloned (Johal and Briggs, 1992), encodes a NADPH-dependent reductase

that is specifically involved in detoxifying HC toxin (Johal and Briggs, 1992). HC toxin is produced by C. *carbonum* race 1 (CCR1), the cause of leaf blight and ear mould in maize. Orthologues of *Hm1* are present in the grass family, and in barley such orthologues contribute to non-host resistance to CCR1 (Sindhu *et al.*, 2008). Many different active loss-of-susceptibility mechanisms of resistance are employed against viruses. For example, the tomato *Tm-1* gene product, which confers resistance to Tomato mosaic virus inhibits the replication of Tomato mosaic virus RNA by binding to replication proteins (Ishibashi *et al.*, 2014). In addition, the allelic tomato *Ty-1*/*Ty-3* resistance genes encode for γ-clade RNA-directed RNA polymerases, which induce RNA-directed DNA methylation to provide immunity against single strand DNA geminiviruses (Butterbach *et al.*, 2014; Verlaan *et al.*, 2013).

Mechanism 8: Passive Loss of Susceptibility—Losing Interaction with Key Host Targets

Loss of interaction of a key host susceptibility factor with a pathogen effector is a common mechanism underpinning recessive *R* genes. Indeed, half of the known *R* genes against viruses involve such a loss of interaction (Truniger and Aranda, 2009). The majority of the so far identified recessive *R* genes act against potyviruses (Kang *et al.*, 2005) and encode mutant translation initiation factors of the 4E or 4G family, which are unable to interact with the cap structure on viral transcripts, thereby conferring resistance to potyviruses (Truniger and Aranda, 2009). In addition, one-third of the *R* genes against *X. oryzae* pv *oryzae*, the causal agent of bacterial blight, are inherited recessively (Q., Liu *et al.*, 2011), and the underlying mechanisms seems often to involve loss of interaction with key host targets. A recessive mutation in the promoter of the rice *xa13* allele does no longer facilitate its manipulation by the TALE AvrXa13, leading to loss of susceptibility (Chu *et al.*, 2006). Likewise, a single amino acid mutation in the rice gamma subunit of transcription factor IIa (OsTFIIAγ5), encoded by the recessive *xa5* resistance gene, confers resistance to various *X. oryzae* strains in

adult plants (Iyer and McCouch, 2004) by restricting bacterial movement (Iyer-Pascuzzi *et al.*, 2008). OsTFIIAγ5 directly interacts with TALEs and is required for their activity (Yuan *et al.*, 2016). The *xa5* mutation likely affects interaction with TALEs to varying degrees, thereby conferring varying degrees of resistance to different *X. oryzae* pathovars (Gu *et al.*, 2009; Huang *et al.*, 2016; Yuan *et al.*, 2016).

Mechanism 9: Passive Loss of Susceptibility by Host Reprogramming

Host reprogramming by mutations in components of cellular pathways is a common strategy leading to durable resistance against a broad range of pathogens. This form of loss of susceptibility is often a recessive trait, but may also involve dominant-negative alleles. The genes involved in this category are often called Adult Plant Resistance (*APR*) genes, as the resistance they confer is often only effective in later stages of plant life (Ellis *et al.*, 2014). Furthermore, *APR* genes often confer partial resistance against a broad range of pathogens (Ellis *et al.*, 2014). It is important to note that the mechanisms underlying this type of resistance do not necessarily act only in the adult plant; therefore, the term *APR* genes does not fully cover this category. Furthermore, perception-based mechanisms as well as loss-of-interaction mechanisms may only function in the adult plant.

The prime example of a senescence-associated loss-of-susceptibility mechanism is mediated by recessive loss-of-function *mlo* (*Mildew Locus O*) alleles in both monocots and dicots (Hückelhoven and Panstruga, 2011). *MLO* encodes a plant-specific integral membrane protein with an unknown molecular function that acts as a negative regulator of cell death in response to both biotic and abiotic stress (Piffanelli *et al.*, 2002). *MLO* loss-of-function alleles are associated with spontaneous cell death (Büschges *et al.*, 1997). In both barley and Arabidopsis, *MLO* is co-expressed with *PEN1*, *PEN2*, and *PEN3* (or its barley orthologues), which are required for an active response against powdery mildews (Humphry *et al.*, 2010). MLO acts as a negative regulator of the *PEN1*/*PEN2*/*PEN3* pathways (Humphry *et al.*, 2010), and in both

barley and Arabidopsis, these genes are required for *mlo*-dependent immunity (Humphry *et al.*, 2006). Therefore, loss of a general suppressor of cell death can confer specific resistance to powdery mildews in both monocots and dicots by deregulation of the PEN1 and PEN2/PEN3 pathways. This principle has been used to engineer a transgene-free tomato *mlo* mutant resistant to powdery mildews (Nekrasov *et al.*, 2017).

Analogous to the *mlo* mechanism, the recessive resistance toward rice blast is conferred by a loss-of-function mutation in *Pi21*, a gene encoding a HMA domain protein that acts to suppress the plant defence response (Fukuoka *et al.*, 2009). It is interesting to note that HMA domain proteins are crucial virulence targets for different pathogenic fungi and that several sequence-unrelated effectors adopt a similar fold that converge on HMA domain proteins (de Guillen *et al.*, 2015; Maqbool *et al.*, 2015). Therefore, engineering loss-of-function alleles of HMA domain-encoding genes analogous to *pi21* may confer broad resistance to various pathogenic fungi in unrelated plant species. Other genes conferring disease resistance through senescence-associated mechanisms include wheat (*Triticum aestivum*) *Lr34*, which encodes an ABC transporter and confers dominant adult resistance to leaf rust, stripe rust, and powdery mildew (Krattinger *et al.*, 2009). In addition, wheat *WKS1* (syn. *Yr36*) encodes a kinase and START domain-containing protein that confers dominant partial resistance to a variety of stripe rusts (Fu *et al.*, 2009). WKS1 functions by reducing the ability of chloroplast thylakoid ascorbate peroxidases (tAPXs) to detoxify peroxides, which contributes to cell death (Gou *et al.*, 2015). Finally, wheat *Lr67* is a dominant *R* gene that confers partial adult resistance against all races of leaf rust, stripe rust, stem rust, and powdery mildew and is caused by a mutation in a predicted hexose transporter that differs by two amino acids from the susceptible allele (Moore *et al.*, 2015). Lr67 displays a dominant-negative effect through heterodimerization with the susceptible allele product, reducing glucose uptake (Moore *et al.*, 2015), ultimately leading to leaf tip necrosis (Herrera-Foessel *et al.*, 2014). The resistance these genes confer is seemingly

dependent on a deregulated initial immune response, resulting in a quicker and stronger immune response that is able to partially suppress the pathogen. However, it is possible that some of these genes in fact act through a different molecular mechanism.

DISCUSSION

Since the cloning and molecular characterization of the first *R* genes in the early 1990s, a steady amount of *R* genes and interactors has been identified (**Figure 1A**). Even still, while Phytozome lists ~14 000 NLR-encoding genes, only 191 have been shown to act as *R* genes (Goodstein *et al.*, 2012). With the advance of novel techniques such as SMRT RenSeq and other next-generation sequencing-based techniques (Steuernagel *et al.*, 2016; Witek *et al.*, 2016), the number of annotated *R* genes is expected to increase at a faster pace than before. Understanding the molecular mechanisms underlying these *R* genes, however, will still require careful case-by-case examination. Indeed, of the examined 314 *R* genes, only 128 have a proposed mechanism, based on an even lower number of detailed studies into molecular mechanisms. The underlying mechanisms of the remaining 186 *R* genes is yet unknown. Understanding these molecular mechanisms in addition to discovering more *R* genes is vital, as it allows us to transfer these traits to other species as well as rationally engineer disease resistance, thereby extending the recognition spectra of *R* genes to outside of what is found in nature. For instance, engineering of the AvrPphB cleavage site in PBS1 to that of other pathogen-derived proteases leads to their recognition in an RPS5-dependent manner (Kim *et al.*, 2016a). Likewise, Executor *R* genes have been engineered to increase their recognition spectra (Hummel *et al.*, 2012; Römer *et al.*, 2009; Zeng *et al.*, 2015). In addition to engineering novel *R* genes, high-throughput screens with effectors on non-host plants can identify *R* genes conferring non-host resistance, which may be very durable when deployed in susceptible hosts.

Critically, the distinction between direct and indirect interactions between the pathogen-derived elicitor and the *R* gene product is not as easily defined as it seems, both experimentally

and conceptually. Direct interactions are not necessarily proven by bimolecular fluorescence complementation, fluorescence resonance energy transfer, or co-immunoprecipitations as the detected complex may contain additional host components required for the interaction. Likewise, even interactions in yeast two-hybrid assays do not exclude conserved eukaryotic factors facilitating the interaction. Direct interactions are best proven with purified proteins and by including non-interacting mutant proteins as controls. A direct interaction is confirmed in vivo when these mutations also disrupt the interaction in vivo. These assays are challenging and have often not been performed. Second, the requirement of an additional host protein for the interaction does not exclude that there is a physical interaction between the *R* gene product and the pathogen-derived effector, for instance, when a host protein acts as a molecular glue to enhance the direct interaction. Therefore, caution must be taken in the interpretation of results.

We have frequently discussed the compliance with guard and decoy models in mechanisms **3**, **5**, and **6**. The molecular mechanistic interpretation of the difference between the guard and decoy model boils down to the question of whether a putative decoy has a function in the absence of the *R* gene. However, even when a guard/decoy is shown to have a role in the absence of the *R* gene, this may result from a decoy function, e.g., when several *R* gene products converge on guarding the same decoy. Similarly, a decoy may still have a function in a pathogen interaction in the absence of a perceived effector if multiple independent effectors from that pathogen converge on the same decoy triggering recognition. Therefore, mechanistically it may be better to refer to a guarded protein acting as a decoy when there is no clear virulence role in targeting the guard/decoy in the absence of an *R* gene in a given pathogen interaction. In other words, decoys only function in specific interactions where they act as co-receptors by mimicking the elicitor target.

It is also interesting to point out genes that have not been found among cloned *R* genes. First, signaling components of defence pathways, such as EDS1, NPR1, etc., are not among cloned

R genes. It might be that these genes are not polymorphic in nature as they are central hubs in interconnected signaling networks, and alterations in these signaling components could disrupt multiple pathways. *R* genes therefore seem to be at the edges of signaling pathways. Second, no pathogenesis-related (*PR*) genes have been found to underlie *R* genes. This could be because most *PR* genes more likely function in immunity in an additive and quantitative manner. Finally, no *R* gene encoding for DAMPs or their receptors have been identified. DAMPs are endogenous "danger" signals that are produced or released upon cellular stress, cellular damage, or non-physiological cell death and can be involved in either triggering or reinforcing immune signaling (Yamada *et al.*, 2016). Because pathogens can cause DAMP release, DAMPs can theoretically facilitate indirect perception of pathogens. It is possible that transfer of DAMPs and their recognition machinery to other species may confer enhanced resistance to specific pathogens, although this remains to be demonstrated.

In conclusion, the past 25 years has revealed the molecular identity of many *R* genes and uncovered the underlying molecular mechanisms governing their functioning. However, much remains to be discovered. Understanding the molecular and structural mechanisms involved in *R* gene mediated disease resistance will allow us to engineer resistance for the crops of the future.

Supplemental Data

Supplemental Data Set 1. List of *R* genes.

Acknowledgements. This work was supported by the European Research Council (616449), by the Clarendon Fund, and by the University of Oxford.

Received July 24, 2017; revised December 14, 2017; accepted January 29, 2018; published January 30, 2018

Chapter 3

Evolution of a guarded decoy protease and its receptor in Solanaceae

Kourelis, J., Malik, S., Mattinson, O., Krauter, S., Kahlon, P.S., Paulus, J.K., van der Hoorn, R.A.L. (2018) Evolution of a guarded decoy protease and its receptor in Solanaceae. *In preparation*

Authors

Jiorgos Kourelis, Shivani Malik, Oliver Mattinson, Sonja Krauter, Parvinderdeep S. Kahlon, Judith K. Paulus, Renier A. L. van der Hoorn

Author contributions

JK conceived the research with feedback from RALvdH. Investigation by JK, OM, SK, and PSK under supervision of JK. SM contributed data for figure 4c. JK wrote the manuscript with feedback from RALvdH.

Justification

In this chapter I describe the evolution of the Cf-2/Rcr3 perception mechanism. I show that *Rcr3* is an ancient gene, whereas *Cf-2* is a novel gene which evolved within an older gene family. I argue that the evolution of *Cf-2* gave a novel molecular role to *Rcr3*, in the recognition of Avr2.

Evolution of a guarded decoy protease and its receptor in Solanaceae

Jiorgos Kourelis[1], Shivani Malik[1], Oliver Mattinson[1], Sonja Krauter[1], Parvinderdeep S. Kahlon[1,2], Judith K. Paulus[1], Renier A. L. van der Hoorn[1*]

[1]Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, South Parks Road, OX1 3RB Oxford, UK

[2]Current address: Lehrstuhl für Phytopathologie, Technische Universität München, Arcisstraße 21, 80333 München, Germany

[*]Correspondence: renier.vanderhoorn@plants.ox.ac.uk (R.A.L. van der Hoorn).

Abstract:

***Cf-2*-dependent recognition of the *Cladosporium fulvum* effector Avr2 requires the papain-like cysteine protease Rcr3. It has been proposed that Rcr3 acts as a decoy for the recognition of Avr2, and that the Rcr3-related Pip1 is the operative target of Avr2. Here we show that transient co-expression of Cf-2, Rcr3, and Avr2 triggers a HR in *Nicotiana benthamiana* and tobacco. *N. benthamiana* is a natural knock-out of *Rcr3* and *Pip1*, and both genes can be resurrected with minimal changes to encode for active proteases. Unlike in *N. benthamiana*, Rcr3 is an active protease in other Solanaceae species. We find that these Rcr3 homologs can be inhibited by Avr2, but this does not necessarily result in a HR. Instead, the ability of these Rcr3 homologs to induce a HR upon co-expression with Cf-2 and Avr2 correlates with the phylogenetic distance to tomato. Protease sensitivity to Avr2 is determined by three residues, and the *Rcr3* homolog that co-evolved with *Cf-2* is not optimized for Avr2 inhibition. Finally, we find that *Cf-2* evolved recently in a tomato-specific subclade of the *Hcr2* gene family. Combined this shows that *Cf-2* evolved recently and co-opted**

Rcr3 **for the recognition of Avr2. This highlights how genetically unlinked indirect recognition mechanism initially evolve and can aid in the application and rational engineering of such mechanisms.**

Keywords: Solanaceae // Plant Innate Immunity // Rcr3 // Pip1 // Avr2 // Cladosporium fulvum // Nicotiana benthamiana

Introduction:

Plants recognize pathogens using both cell-surface and intra-cellular receptors. Many of these receptors are encoded by polymorphic resistance (*R*) genes. The products of these *R* genes function either by directly recognizing pathogen components, or by indirectly recognizing the interaction of pathogen components with host components or modification of these host components by pathogen components (Kourelis and van der Hoorn, 2018). This second, indirect mechanism is also referred to as the "guard" or "decoy" model, in which a host component is either monitored – "guarded" - by a *R* gene product or in which a host component mimics an operative target of the pathogen, trapping it into a recognition event through a corresponding receptor (van der Hoorn and Kamoun, 2008).

In plants, receptors encoded by *R* genes typically belong to one of two classes of receptors: i) cell-surface receptor-like proteins (RLPs) and receptor-like kinases (RLKs), and ii) intracellular Nod-like receptors (NLRs). NLRs can contain additional integrated domains (IDs), some of which were shown to interact with pathogen elicitors, thereby triggering immune responses. One clade of ancient grass NLR-IDs evolve in syntenic genomic loci by repeated integrations of different IDs, marked by a 43 amino acid integration motif (Bailey *et al.*, 2018). How a guard/decoy mechanism evolves in genetically unlinked receptor/pathogen target, however, is not well understood. One well-studied example of a genetically unlinked

guard/decoy mechanism is the recognition of the tomato leaf mould causing pathogen *Cladosporium fulvum* (syn. *Passalora fulva*) by the product of the tomato *R* gene *Cf-2*.

Cf-2 was introgressed into cultivated tomato (*Solanum lycopersicum*) from the wild currant tomato (*Solanum pimpinellifolium*) and encodes for a RLP with extracellular leucine-rich-repeats (LRRs) (Dixon *et al.*, 1996). *Cf-2* recognizes *C. fulvum* strains producing a secreted small cysteine-rich effector called Avr2 (Dixon *et al.*, 1996; Luderer *et al.*, 2002). This recognition also requires *Rcr3*, which encodes for a papain-like cysteine protease (PLCP) (Dixon *et al.*, 2000). Avr2 acts as a protease inhibitor, and it is the interaction of Avr2 with Rcr3 which is recognized by *Cf-2* resulting in a hypersensitive response (HR) and immunity to *C. fulvum* (Rooney *et al.*, 2005).

Besides Rcr3, Avr2 also inhibits the Rcr3-related immune protease Pip1 (Shabab *et al.*, 2008). Both *Rcr3* and *Pip1* are regulated as pathogenesis-related (*PR*) genes (Shabab *et al.*, 2008), but Pip1 is more abundant in the apoplast than Rcr3 (Shabab *et al.*, 2008). While knock-out of *Rcr3* does not result in enhanced susceptibility to *C. fulvum* in tomato lines lacking *Cf-2*, knock-down of *Pip1* strongly increases susceptibility to *C. fulvum* (Ilyas *et al.*, 2015). This indicates that in this interaction Pip1 is the actual operative target of Avr2, while Rcr3 acts as a decoy to trap the pathogen into a recognition event (van der Hoorn and Kamoun, 2008). This hypothesis leads to two potential evolutionary trajectories: i) Rcr3 is a relatively novel gene, which has evolved to maximize the interaction with Avr2, or ii) Rcr3 is more generally required for the function of different immune receptors and specifically evolves to interact more strongly with different pathogen components. Alternatively, iii) if Rcr3 is guarded by Cf-2 we would expect *Cf-2* to have evolved to use pre-existing machinery. In this scenario Rcr3 does not necessarily need to be the main target of Avr2.

Here we investigated the evolution of the Cf-2/Rcr3 perception system in Solanaceae to determine which gene likely evolved first. We also established transient HR assays to study Avr2/Rcr3/Cf-2 by agroinfiltration in *N. benthamiana* and studied the role of *Rcr3* and *Pip1* in *N. benthamiana*. This work provides exciting new insights into the evolution of a guarded decoy and its receptor in Solanaceae.

Results:

Co- expression of Cf-2, Rcr3, and Avr2 triggers a robust HR in N. benthamiana

We first established a transient HR assay for Cf-2-dependent recognition of Avr2 in *N. benthamiana*. Co-infiltration of Agrobacterium strains containing binary vectors for the expression of Rcr3, Avr2, and a synthetic *Cf-2* construct in which the repetitive repeats present in the natural DNA sequence were removed triggers a HR which starts at 2-3 days-post-infiltration (dpi) and is fully developed within 5dpi (**Figure 1a**). No HR was induced when any of the three components were omitted (**Figure 1a**).

Transient expression of Cf-2, Rcr3^{pim} and Avr2 by agroinfiltration in tobacco (*Nicotiana tabacum* cv. *Petite Havana SR1*) induces a HR (**Figure 1b**), which is also absent when any of the three components is omitted (**Figure 1b**). These data show that HR signalling downstream of Cf-2/Rcr3 is conserved in *Nicotiana* species, but that these *Nicotiana* species do not already carry a functional *Cf-2* or *Rcr3*. These data indicate that either Rcr3 is absent in *Nicotiana* species, consistent with the hypothesis that *Rcr3* evolved only recently, or that the *Nicotiana* Rcr3 is not able to participate in Cf-2-dependent Avr2 recognition.

We next replaced each of the three components to study the specificity of the transient HR assay. We first replaced Avr2 with a previously described Avr2 mutant which lacks the last six residues (Avr2Δ6) and which cannot inhibit Rcr3 nor trigger HR in tomato lines containing *Cf-2* (van't Klooster *et al.*, 2011). Indeed, the combination with the Avr2Δ6 mutant did not

trigger a HR (**Figure 1c**). We also replaced Avr2 with the only known natural variant of Avr2 carrying a leucine to valine (L43V) substitution (Stergiopoulos *et al.*, 2007). Avr2(L43V) still induces a HR upon co-expression with Cf-2 and $Rcr3^{pim}$ (**Figure 1c**). Finally, *Cf-2* has also been reported to recognize the nematode *Globodera rostochiensis* effector GrVAP1 in an *Rcr3*-dependent manner (Lozano-Torres *et al.*, 2012). However, co-expression of GrVAP1 with Cf-2 and $Rcr3^{pim}$ in *N. benthamiana* by agroinfiltration did not trigger a HR (**Figure 1d**). We therefore omitted GrVAP1 from subsequent assays.

Likewise, we replaced $Rcr3^{pim}$ with the Rcr3-related PLCP Pip1, which cannot participate in Cf-2-dependent Avr2 recognition. Similarly, this did not trigger a HR (**Figure 1e**). Next, we investigated *Cf-2*-dependent induction of auto-necrosis. The *S. lycopersicum* allele or *Rcr3* ($Rcr3^{lyc}$) induces Avr2-independent auto-necrosis in tomato lines carrying *Cf-2* (Krüger *et al.*, 2002). To determine whether we could also induce auto-necrosis in our transient assay we co-expressed Cf-2 and $Rcr3^{lyc}$ by agroinfiltration. However, we were unable to detect the development of HR or necrosis (**Figure 1f**). This indicates that either our assay is not sensitive enough to detect auto-necrosis or an additional component is absent in this assay.

Finally, we replaced Cf-2 by Cf-5. *Cf-5* is allelic to *Cf-2* and provides resistance against *C. fulvum* races producing Avr5 (Dixon *et al.*, 1998). Like *Avr2*, *Avr5* encodes for a small secreted protein with several putative disulphide bridges and no significant sequence homology to proteins with known functions (Mesarich *et al.*, 2014). Replacement of Cf-2 by Cf-5 did not trigger a HR with Avr2 and Rcr3, showing that Rcr3/Avr2-induced HR is *Cf-2*-specific (**Figure 1g**). Co-expression of Cf-5 with Avr5 triggers a HR, showing that this HR does not require co-expression with Rcr3 (**Figure 1g**), consistent with previous observations that *Cf-5* does not require *Rcr3* for the recognition of *C. fulvum* races producing Avr5 (Dixon *et al.*, 2000). Cf-2/Rcr3 does not trigger a HR with Avr5 (**Figure 1g**), consistent with the classic gene-for-gene model.

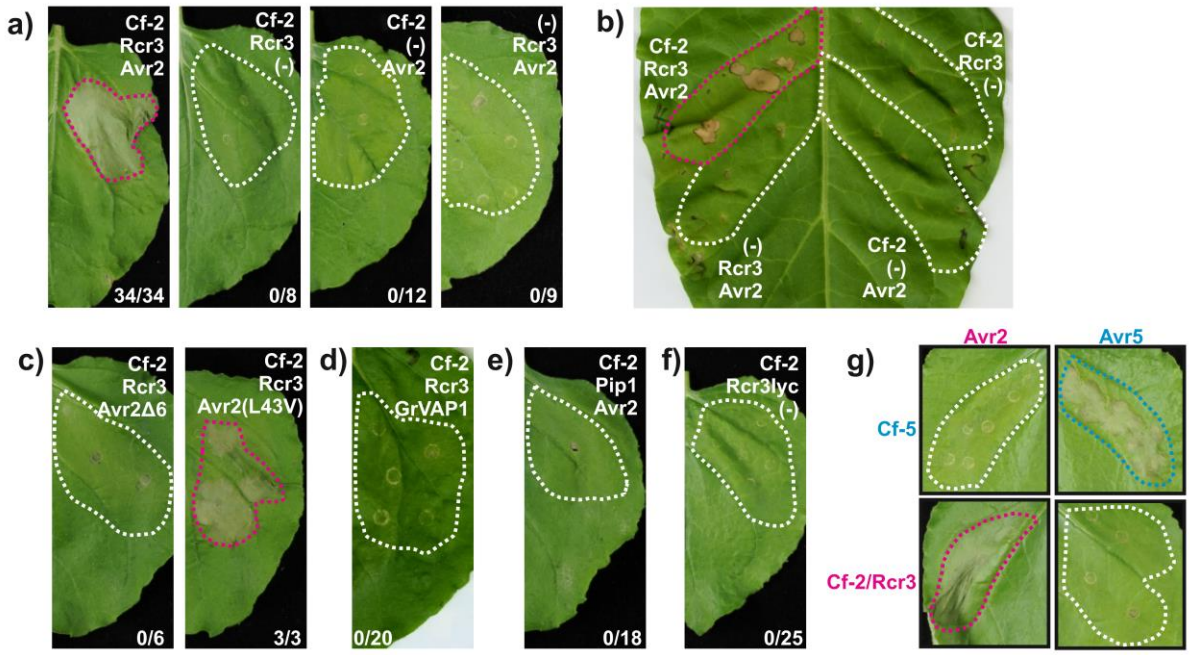


Figure 1: Transient co-expression of Cf-2, Rcr3, and Avr2 triggers a HR in *Nicotiana benthamiana and tobacco*. **a)** Co-infiltration of *A. tumefaciens* containing binary vectors for the expression of a codon-optimized Cf-2, Rcr3 and/or Avr2. A HR is triggered in *N. benthamiana* only when all three components are present. **b)** Transient expression of Cf-2, Rcr3, and Avr2 without the P19 silencing inhibitor in *N. tabacum* cv. *Petite Havana SR1* by agroinfiltration. A HR is triggered only when all three components are present. **c)** Transient co-expression of Cf-2 and Rcr3 with either an Avr2 mutant (Avr2Δ6) or with a natural variant of Avr2 containing the L43V substitution (Avr2(L43V)). Avr2Δ6 fails to trigger a HR whereas Avr2(L43V) triggers a HR. **d)** Transient co-expression of Cf-2, Rcr3, and GrVAP1 does not trigger a HR in *N. benthamiana*. **e)** Transient co-expression of Cf-2 and Avr2 with Pip1 does not trigger a HR. **f)** Transient co-expression of Cf-2 with the autonecrosis-inducing *S. lycopersicum* allele of Rcr3 ($Rcr3^{lyc}$) does not trigger a HR in *N. benthamiana*. **f) g)** Transient expression of codon-optimized Cf-5 with Avr5 triggers a HR. **a-g)** *A. tumefaciens* containing different binary vectors were combined at a final OD_{600} = 0.25 each in a 1:1:1 ratio. The silencing inhibitor P19 was present in the backbone of all binary vectors except for those used in tobacco. All pictures taken 5 dpi. Numbers indicate the number of leaves showing HR.

N. benthamiana Rcr3 is a pseudogene and can be resurrected, but it cannot participate in Cf-2-dependent Avr2 recognition

Transient Cf-2-dependent Avr2 recognition in *N. benthamiana* requires co-expression with Rcr3. We therefore investigated whether this is because *N. benthamiana* lacks a functional Rcr3, or whether *N. benthamiana* Rcr3 is unable to interact with either Avr2 or Cf-2. In the previously published *N. benthamiana* draft genome (Bombarely *et al.*, 2012) we identified a

contig containing both an *Rcr3* and *Pip1* homolog (*NbRcr3a* and *NbPip1*) (**Figure 2a**). This locus is syntenic with the *Rcr3*/*Pip1* locus in tomato, potato, and pepper (Ilyas *et al.*, 2015). In addition to NbRcr3a and NbPip1, this syntenic block also contains homologs of the *Pip1-flanking protease 1* (*NbPfp1*) and *Rcr3-flanking protease 1* (*NbRfp1*), as well as a homolog of lipase genes (**Figure 2a**). Additionally, we identified a second *N. benthamiana Rcr3* homolog present on a different contig (*NbRcr3b*). Since this contig also contained several lipase genes, it is possible that these *NbRcr3* homologs reside near each other on the same chromosome. Alternatively, they may reside on different chromosomal location explained by the ancient allotetraploid origin of *N. benthamiana*.

A closer look revealed that both *NbRcr3a* and *NbRcr3b* are pseudogenized. Mutations in *NbRcr3a* cause frameshifts resulting in truncated proteins. These frame-shift mutations appear to have independently arisen in the *N. benthamiana* NWA, NT, and LAB accessions (**Figure 2b**). We could not identify *NbRcr3a* in the published *de novo* transcriptomes of the *N. benthamiana* SA, WA, or QLD accessions, nor could we identify *NbRcr3b* sequences containing a start-codon.

We corrected the frameshifts in $NbRcr3a^{LAB}$ and produced this mutant ($NbRcr3_{CF}$) by agroinfiltration. Activity-based protein profiling with the MV201 probe, a fluorescent derivative of the general PLCP inhibitor E-64 (Richau *et al.*, 2012), was used to test the activity of $NbRcr3_{CF}$. However, $NbRcr3_{CF}$ is not an active protease (**Figure 2c**). Upon aligning the $NbRcr3_{CF}$ protein with all Arabidopsis and tomato PLCPs we realized there were two additional mutations in otherwise highly conserved residues: a glutamic acid into alanine substitution (E213A) and a tyrosine into histidine substitution (Y216H). Correcting H216Y is sufficient to restore activity to $NbRcr3_{CF}$ ($NbRcr3_{CF/H216Y}$) (**Figure 2c**). While the additional mutation A213E is not required for restoring catalytic activity ($NbRcr3_{CF/A213E/H216Y}$, from here onwards referred to as $NbRcr3_{res}$, **Figure 2b, c**), it may be required for normal physiological

function. The E213A substitution is shared between the different *N. benthamiana* accessions (**Figure 2b**) and may have arisen first, with subsequent frame-shift mutations and non-synonymous mutations occurring later in different *N. benthamiana* populations.

Although NbRcr3$_{res}$ is an active protease, it is not inhibited by Avr2 (**Figure 2c**). Consistent with its lack of Avr2 inhibition, NbRcr3$_{res}$ cannot trigger HR when co-expressed with Cf-2 and Avr2 in *N. benthamiana* (**Figure 2d**). We suspected this was because NbRcr3$_{res}$ carries a glycine at position 194, because previously it was shown that a naturally occurring allele of Rcr3 carrying the N194D substitution was not inhibited by Avr2 (Shabab *et al.*, 2008). Indeed, the NbRcr3$_{res}$ G194N mutant is efficiently inhibited by Avr2 (**Figure 2c**), indicating that N194 enhances the interaction between Rcr3 and Avr2. However, co-expression of the NbRcr3$_{res}$ G194N mutant with Cf-2 and Avr2 does not trigger HR in *N. benthamiana* (**Figure 2d**), indicating that NbRcr3$_{res}$ is unable to interact with *Cf-2*.

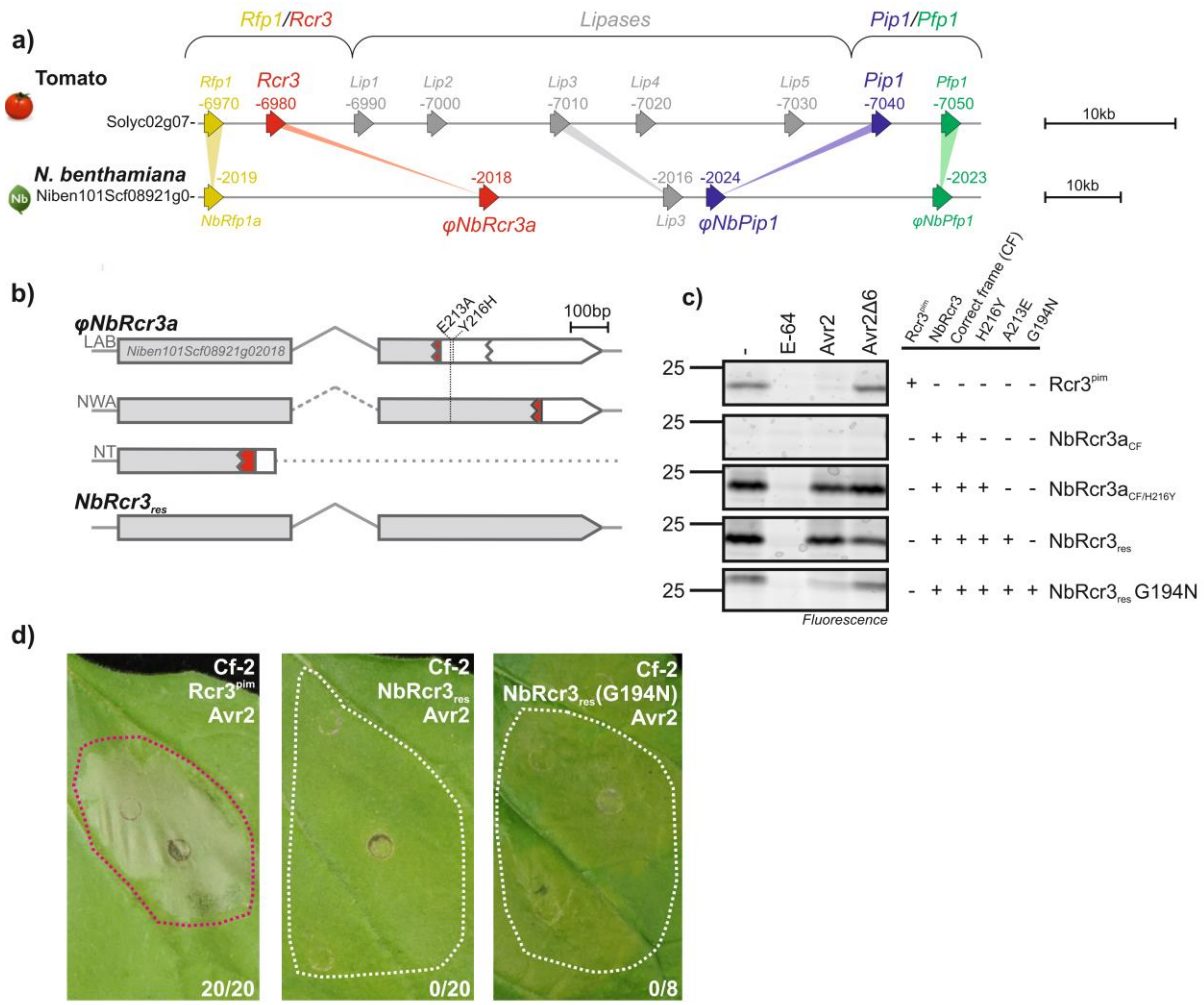


Figure 2: *Nicotiana benthamiana* is a natural knock-out of Rcr3. **a)** Synteny of the *N. benthamiana* and tomato *Rcr3*/*Pip1* locus. **b)** Schematic representation of mutations in *NbRcr3a* in different *N. benthamiana* accessions. Dotted line indicates unknown sequence, red indicates different ORF after indels, white indicates the untranslated region corresponding to *Rcr3*. Additional mutations leading to an inactive protease are indicated. **c)** Activity-based protein-profiling of Papain-like cysteine proteases. $Rcr3^{pim}$ and NbRcr3a mutants ($NbRcr3_{CF}$, $NbRcr3_{CF/H216Y}$, $NbRcr3_{res}$, and $NbRcr3_{res}$ G194N) were co-expressed with silencing inhibitor P19 by agroinfiltration of *N. benthamiana* leaves. 2 dpi apoplastic fluids were isolated and pre-incubated for 45 min with and without 100 µM E-64, 2 µM Avr2 or Avr2Δ6, and labelled for 3 hours with 2 µM MV201. Samples were separated on SDS-PAGE gels and scanned for fluorescence. **d)** Transient expression of $NbRcr3_{res}$ or the $NbRcr3_{res}$ G194N mutant with Cf-2 and Avr2 does not trigger a HR in *N. benthamiana*. Binary vectors contained the silencing inhibitor P19 in the backbone, and the *A. tumefaciens* cultures were combined in a final OD_{600} = 0.25 each in 1:1:1 ratio. Pictures were taken at 5 dpi. Numbers indicate the number of leaves showing HR.

Pip1: an important immune protease in tomato

Like *Rcr3*, *Pip1* is highly conserved in Solanaceae (**Figure S1**). *Pip1* acts as an important immune protease in tomato, providing basal resistance against bacterial (*Pseudomonas syringae* pv. *tomato* DC3000), fungal (*Cladosporium fulvum*), and oomycete (*Phytophthora infestans*) pathogens (Ilyas *et al.*, 2015). The mechanism by which *Pip1* provides basal resistance against these diverse pathogens is unknown. Pip1 activity is also strongly increased prior to Cf-4/Avr4-dependent HR (Sueldo *et al.*, 2014), and silencing of *Pip1* in *N. benthamiana* blocked HR induced by Cf-4/Avr4 (Xu *et al.*, 2012). We therefore tested whether Pip1, like Rcr3, may be required for RLP-dependent signalling. We used CRISPR/Cas9 genome editing (Nekrasov *et al.*, 2013) to generate a mutant line carrying frame-shifts in *NbPip1* (*pip1-1*) (**Figure 3a**). Surprisingly, the *N. benthamiana pip1-1* line was still able to trigger HR upon co-expression of Cf-4 and Avr4 (**Figure 3b**). Additionally, the *pip1-1* line is not more susceptible to *Pseudomonas syringae* pv. *tomato* DC3000 ΔhopQ1-1 (*Pst*DC3000ΔhopQ) (**Figure 3c,d**), unlike antisense *Pip1* tomato lines (Ilyas *et al.*, 2015). Also *TRV::NbPip1 N. benthamiana* plants are not more susceptible to *Pst*DC3000ΔhopQ (**Figure 3d**).

We therefore took a closer look at the *NbPip1* gene in *N. benthamiana*. We realized that two deletions of 2bp and 24bp respectively in the 3' end of the *NbPip1* gene cause a frame-shift that replaces the last 13 amino acids of NbPip1, with a different sequence of 20 amino acids (**Figure 3e**). An important cysteine required to make a disulphide-bridge critical for PLCP function (Richau *et al.*, 2012) is present in this region. Two cysteines are present in the 20 amino acid replacement sequence, but these cysteines are not in the same position or sequence context as the cysteine in the original sequence, and are therefore likely not able to form a disulphide bridge. These deletions are likely very recent, as they are only present in the closely related

LAB and SA accessions, but not in the more distantly related NWA, WA, NT, and QLD accessions (**Figure 3e**). Activity-based protein profiling shows that NbPip1 is indeed not an active PLCP when produced by agroinfiltration (**Figure 3f**). Correction of the 3' deletion results in an active resurrected PLCP ($NbPip1_{res}$), but this protease is not inhibited by Avr2, unlike tomato Pip1 (**Figure 3f**).

Because both *Rcr3* and *Pip1* are pseudogenized in *N. benthamiana*, we analysed the *PLCP* gene family in the *N. benthamiana* genome for additional pseudogenization. We realized that almost as many putative PLCP pseudogenes as genuine PLCP genes are found in *N. benthamiana*, similar to what is described for the *N. benthamiana* subtilisin gene family (Kourelis *et al.*, 2018) (Chapter 7 this thesis) (**Figure 3g**, **Figure S2**, **Figure S3**, **Table S2**). The general pseudogenization may be the result of a contraction of the functional genome after polyploidization event approximately 6 million years ago, giving rise to the *Nicotiana* section to which *N. benthamiana* belongs (Clarkson *et al.*, 2017).

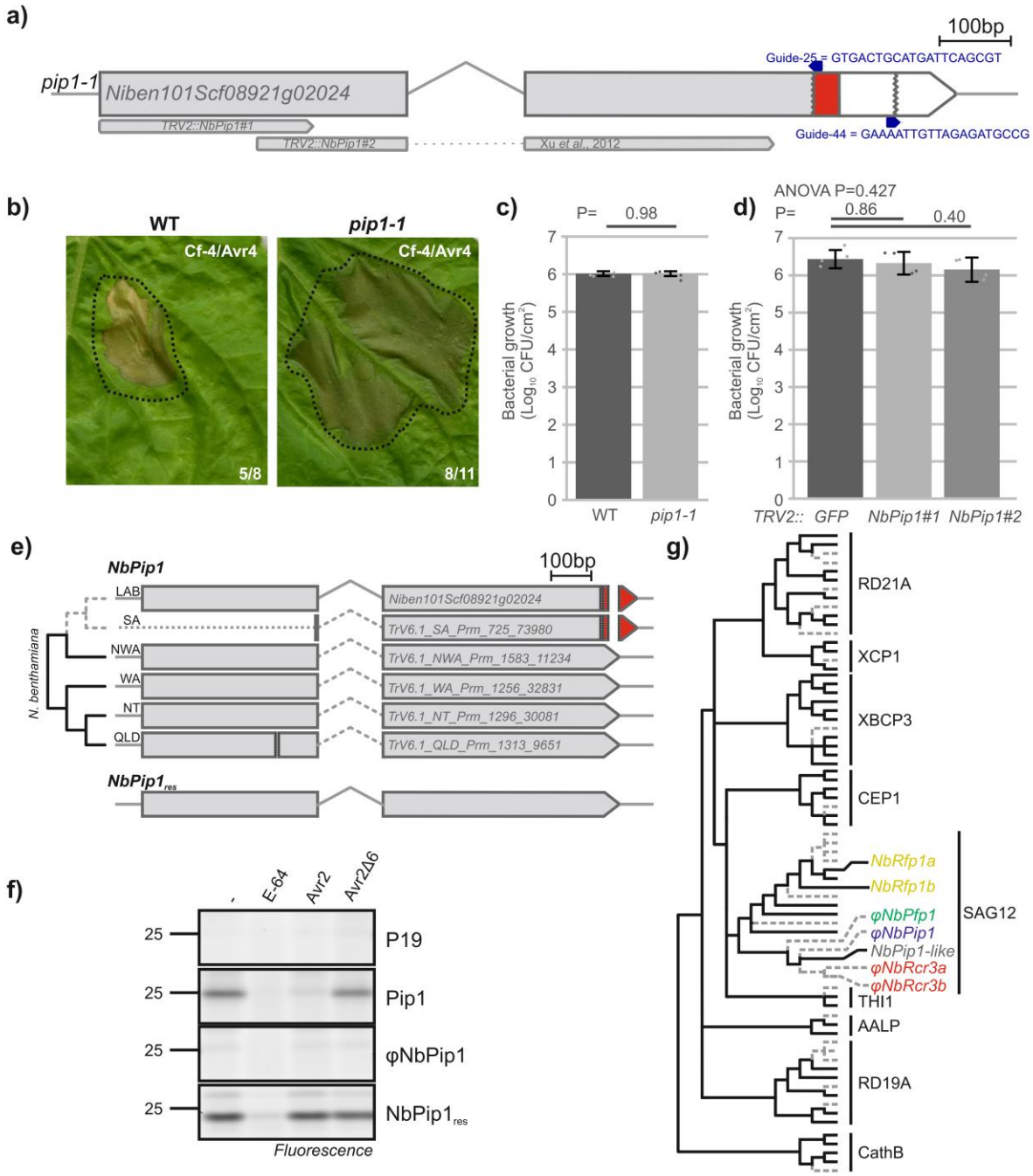


Figure 3: *NbPip1* is also a pseudogene and it is not required for immunity or HR signalling. **a)** Schematic representation of the *NbPip1* gene-model. CRISPR/Cas9 frame-shift mutant of *NbPip1* (*pip1-1*) causes a frameshift leading to an alternative open reading frame (ORF) after a 1bp deletion (indicated in red). White is the untranslated region of *NbPip1*. Two fragments of *NbPip1* used for VIGS are indicated below the sequence. **b)** Transient co-expression of Cf-4 and Avr4 in both wild-type and *pip1-1* mutant *N. benthamiana* plants triggers a HR. Cf-4 and Avr4 binary vectors contained the silencing inhibitor P19 in the backbone. *A. tumefaciens* cultures were combined to a final OD_{600} = 0.25 each in a 1:1 ratio. Pictures were taken 5 dpi. Numbers indicate the number of leaves showing HR. **c)** The *pip1-1* mutant line is not more susceptible to *P. syringae* pv. *tomato* DC3000 ΔhopQ1-1 (*Pst*DC3000ΔhopQ). Plants were infiltrated with *Pst*DC3000ΔhopQ at an OD_{600} = 0.0002 (1 x 10^5 CFU/ml), and at 3 dpi bacterial growth was measured by colony counting. Error bars represent standard deviation of n=4 biological replicates (indicated with jittered dots) at 72 hrs

Figure 3. (continued). post-infection. **d)** Virus-induced gene-silencing targeting *NbPip1* using two independent fragments (*TRV2::NbPip1#1* and *TRV2::NbPip1#2*) does not affect susceptibility to *Pst*DC3000ΔhopQ. Infection done as in (**c**). *TRV2::GFP* is included as a negative control. Error bars represent standard deviation of n=4 biological replicates (indicated with jittered dots) at 72 hrs post-infection. **e)** A 2 bp and 24 bp deletion present in the 3'end of *NbPip1* gene of the *N. benthamiana* LAB and SA accessions, but not the NT, QLD, WA, and NWA accessions, causes an inactive *NbPip1*. Dotted line indicates unknown sequence, red indicates different ORF. **f)** Activity-based protein-profiling of Papain-like cysteine proteases. Pip1, φNbPip1, and an NbPip1 mutant with a corrected 3' end ($NbPip1_{res}$) were co-expressed with the silencing inhibitor P19 by agroinfiltration of *N. benthamiana* leaves. Apoplastic fluids were isolated at 4 dpi, pre-incubated for 45 min with or without 100 µM E-64, 2 µM Avr2 or Avr2Δ6, or DMSO, and labelled for 3 hours with 2 µM MV201. Samples were separated on SDS-PAGE gels, scanned for fluorescence. P19 alone served as a negative control. **g)** The evolutionary history of the *N. benthamiana PLCP* gene family was inferred using the Maximum Likelihood method based on the Whelan and Goldman model (Whelan and Goldman, 2001). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed (Felsenstein, 1985). Putative pseudogenes are indicated with grey dotted lines. The PLCP clades were named according to (Richau *et al.*, 2012). See **Figure S2** for all names.

Rcr3 is conserved across the Solanaceae family

We next investigated whether Rcr3 homologs from other species could participate in Cf-2-dependent Avr2 recognition. We identified the *Rcr3*/*Pip1* locus in a syntenic block in tomato species (*Solanum lycopersicum* and *S. pennellii*), Peruvian wolfpeach (*Solanum lycopersicoides*), potato (*Solanum tuberosum*), eggplant (*Solanum melongena*), pepper species (*Capsicum annuum glabriusculum*, *C. annuum zunla-1*, *C. annuum* cv CM334), and *Nicotiana tabacum* (**Table S1**). In addition to *Rcr3* and *Pip1*, we identified *Pfp1* and *Rfp1* on the same locus in most plant species (**Table S1**). *Rcr3*, *Pip1*, and *Pfp1* cluster together in one subclade of the SAG12 PLCP subfamily (**Figure S3**). One other PLCP, which we designated *Pip1-like*, is also conserved in Solanaceae and clusters together with *Rcr3*, *Pip1*, and *Pfp1* but resides on a different genomic locus (**Figure S1, Figure S3**). *Rfp1* is more distantly related to *Rcr3*, *Pip1*, and *Pfp1*, and clusters together with SAG12 subfamily PLCPs present of different chromosomes (**Figure S3**). *Rcr3* and *Pip1* are Solanaceae-specific, as we could not identify any clear homologs of *Rcr3* and *Pip1* outside of the Solanaceae family (**Figure S1**). Since *Pip1* and *Rcr3* are present in all Solanaceae species, but not outside of this family, we conclude that

these genes evolved over 50Mya, when the Solanaceae family diverged from other plant families (Wilf *et al.*, 2017).

Identifying true orthologues of *Rcr3* and *Pip1* is complicated because of duplications as well as pseudogenization events. Some of these duplications may be species-specific, as *Rcr3* homologs largely cluster according to species (**Figure 4a, Table S1**).

Activity-based protein profiling of Rcr3 homologs produced by agroinfiltration showed that all 13 tested Rcr3 homologs are active proteases (**Figure 4b**), making *N. benthamiana* the exception having no active *Rcr3*. Importantly all active Rcr3 homologs except for the tested *Nicotiana* Rcr3 homologs can be inhibited by Avr2 (**Figure 4b**). Previously it was shown that the N194D variant of Rcr3 prevents Avr2 inhibition (Shabab *et al.*, 2008). The *Nicotiana* Rcr3 homologs carry a lysine at position 194, and mutating this lysine into asparagine (K194N) results in *Nicotiana* Rcr3 mutants which can be inhibited by Avr2 (**Figure S4**). All tomato and potato Rcr3 homologs can trigger HR when co-expressed with Cf-2 and Avr2 (**Figure 3c**). By contrast, the tested eggplant and pepper Rcr3 homologs only trigger a weak response (**Figure 3c**), whereas the petunia (*Petunia inflata*), and *Nicotiana* Rcr3 homologs do not trigger a HR when co-expressed with Cf-2 and Avr2 (**Figure 3c**), and neither do the *Nicotiana* Rcr3 K194N mutants (**Figure S4**). Therefore, the ability of Rcr3 homologs to participate in Cf-2-dependent Avr2 recognition correlates with the evolutionary distance to tomato. The absence of HR while these Rcr3 homologs can be inhibited by Avr2 indicates their inability to interact with *Cf-2*.

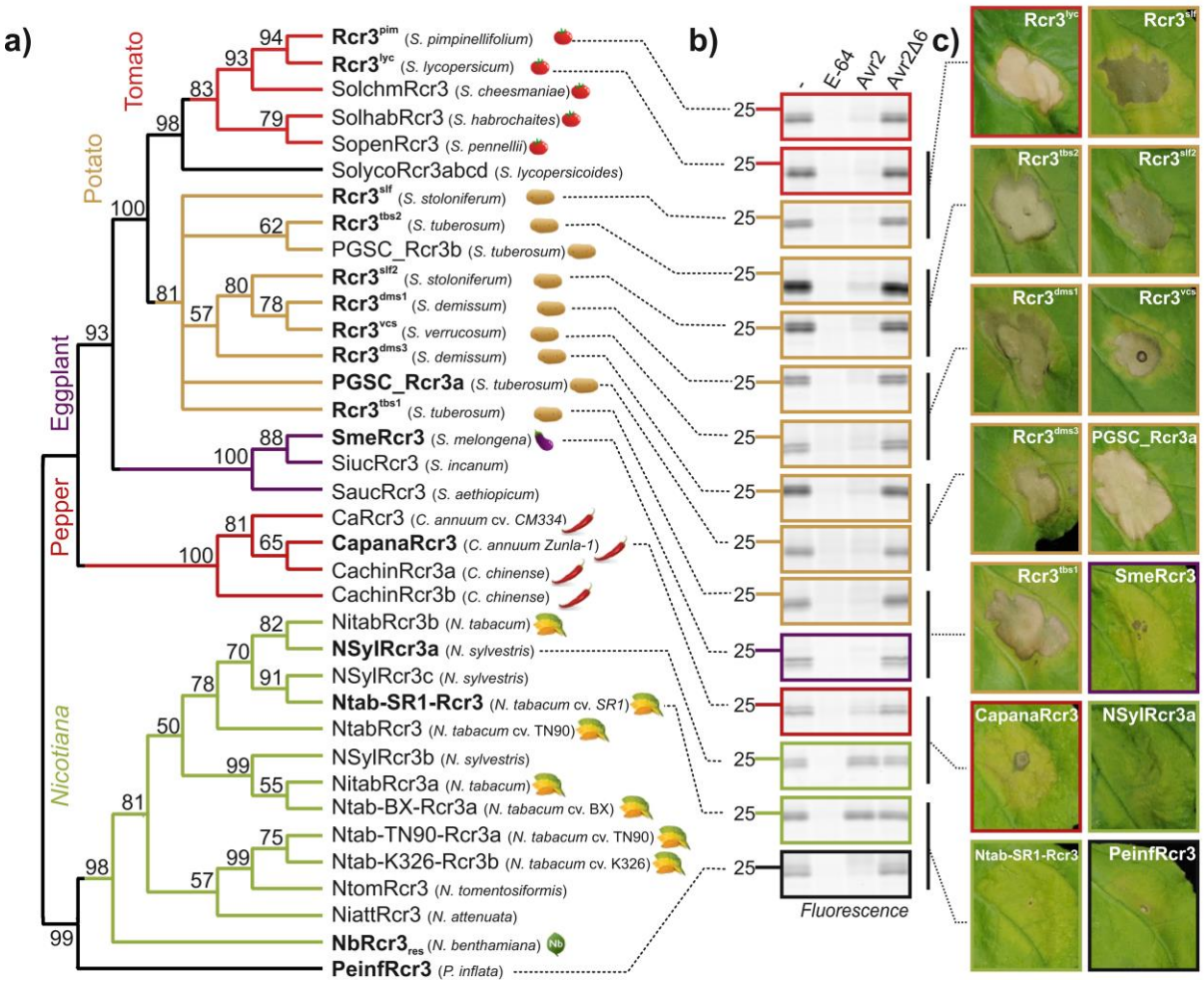


Figure 4: Rcr3 is an active protease in Solanaceae. **a)** The evolutionary history of Rcr3 was inferred using the Maximum Likelihood method based on the Whelan and Goldman model. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history. Shown are percentage bootstrap values. Rcr3 homologs tested for Cf-2-dependent Avr2 recognition and Avr2 inhibition are printed in bold. See **Table S3** for a summary of polymorphic residues in the peptidase C1 domain of the cloned Rcr3 homologs. **b)** Activity-based protein-profiling of Papain-like cysteine proteases. Rcr3 homologs were co-expressed with the silencing inhibitor P19 in *N. benthamiana* leaves by agroinfiltration. Apoplastic fluids were isolated at 2 dpi, pre-incubated for 45 min with or without 100 µM E-64, 4 µM Avr2 or Avr2Δ6, or DMSO, and labelled for 3 hours with 2 µM MV201. Samples were separated on SDS-PAGE gels and scanned for fluorescence. **c)** Co-expression of Cf-2, Avr2 and Rcr3 homologs in *N. benthamiana* by agroinfiltration. Tomato and potato Rcr3 homologs can trigger a HR, while the pepper and eggplant Rcr3 homologs give a weaker response and the Petunia and *Nicotiana* Rcr3 homologs do not trigger any response. Binary vectors contained the silencing inhibitor P19 in the backbone, and the *A. tumefaciens* cultures were diluted to OD_{600} = 0.25 each in 1:1:1 mix. Pictures taken at 8 dpi.

Three residues determine PLCP sensitivity for Avr2

Avr2 appears to inhibit most Rcr3 homologs, with N194 being an important determinant of this interaction. However, we realized that N194 is not the only polymorphic residue involved in the inhibition by Avr2. This is because the Rcr3 homolog from pepper, CapanaRcr3, has an aspartic acid residue at this position (D194), but is still efficiently inhibited by Avr2 (**Table S3**), and similarly, Pip1 has a glutamic acid (E194) at this position, but it also is efficiently inhibited by Avr2 (**Figure 5a**) (Shabab *et al.*, 2008; Kaschani *et al.*, 2010). This indicates that there are variant residues that can compensate for variations in N194.

Some tomato *Rcr3* alleles with the N194D substitution were found to still trigger HR in the presence of *Cf-2* and Avr2. This was associated with a substitution of arginine for glutamine at position 151 (R151Q, (Hörger *et al.*, 2012)). Indeed, CapanaRcr3 has Q151 (**Table S3**), while Pip1 has V151. To test this association, we produced N194D single mutant, and R151Q/N194D and R151V/N194D double mutants of $Rcr3^{pim}$ by agroinfiltration. $Rcr3^{pim}$(N194D) is not inhibited by Avr2 at the tested concentrations (**Figure 5b**), while the R151Q/N194D and R151V/N194D double mutants are slightly better inhibited by Avr2, confirming a role for R151 in the interaction with Avr2 (**Figure 5b**). Similarly, Pip1(V151R) is slightly less inhibited by Avr2, indicating that R151 acts as a repulsive residue (**Figure 5A**).

However, in contrast to $Rcr3^{pim}$(N194D), $Rcr3^{lyc}$(N194D) was still inhibited by Avr2 while both contain R151 (**Figure 5c**). $Rcr3^{pim}$ and $Rcr3^{lyc}$ differ by only six amino acid residues in the mature protease, and within tomato Rcr3 homologs, one of these polymorphisms, Q284R, appears to be specific to $Rcr3^{pim}$ which co-evolved with Cf-2 (Hörger *et al.*, 2012). Pip1 also carries R284. Since this residue resides around the substrate binding groove together with R151 and N194 (**Figure 5d**) we investigated its potential role in Avr2 interaction. Indeed, Pip1(R284Q) (**Figure 5a**) and $Rcr3^{lyc}$(N194D/R284Q) (**Figure 5c**) are less efficiently inhibited

by Avr2, while the opposite occurs for the Rcr3^{pim}(N194D/Q284R) as well as Rcr3^{pim}(R151Q/N194D/Q284R) and Rcr3^{pim}(R151V/N194D/Q284R) (**Figure 5b**).

In 360 previously sequenced tomato genomes (Lin *et al.*, 2014) we could only identify the R284Q variation in Rcr3, and no variation at these positions in any of the other tomato PLCPs (**Figure S5**). This indicates that in tomato species only Rcr3 has polymorphic residues at these positions. Additionally, this indicates Rcr3^{pim} does not contain all the residues for optimal inhibition by Avr2 (**Figure 5e**).

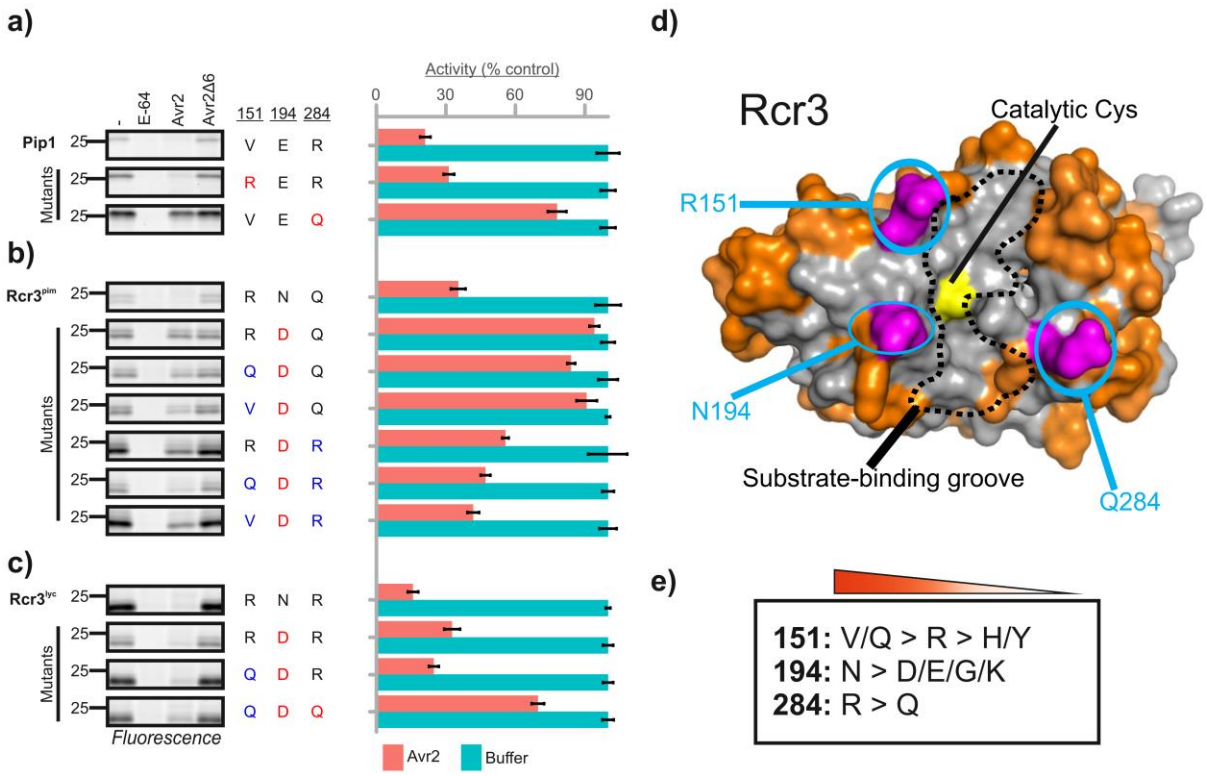


Figure 5: PLCP inhibition by Avr2 is determined by three residues. **a)** Pip1, **b)** Rcr3pim, **c)** Rcr3lyc mutants were co-expressed with the silencing inhibitor P19 by agroinfiltration of *N. benthamiana* leaves. Apoplastic fluids were isolated at 2dpi, pre-incubated for 45min with or without 100µM E-64, 4µM Avr2 or Avr2Δ6, or DMSO, and labelled for 3 hours with 2µM MV201. Samples were separated on SDS-PAGE gels and scanned for fluorescence. Gels were quantified in ImageJ (N=3 DMSO control, N=6 Avr2 inhibition). **d)** Homology model of Rcr3^{lyc} based on 1s4v (Ilyas *et al.*, 2015). Dotted line marks the substrate-binding groove, the catalytic cysteine is indicated in yellow, orange indicates polymorphic residues between Rcr3 and Pip1. Magenta indicates the R151, N194, and Q284 residues. **d)** Summary of residues promoting Avr2 inhibition in descending order of affinity.

Cf-2 belongs to a new-world specific clade of the Hcr2 gene family

In contrast to *Rcr3*, *Cf-2* appears to be an evolutionary novel gene which has only recently evolved by intragenic recombination (Dixon *et al.*, 1996). This intragenic recombination results in a highly repetitive gene structure evident both on a nucleotide and amino acid level (Dixon *et al.*, 1996).

To investigate the evolutionary origins of *Cf-2* we constructed a phylogenetic tree of the *Cf-2* homologs based on the amino acids encoded by the region corresponding to the conserved 5' and 3' region. This phylogeny indicates that *Cf-2* homologs are conserved in Solanaceae (**Figure 6a**). *Cf-2* homologs contain six distinct regions termed domain A to F (Dixon *et al.*, 1996). The N-terminal region containing the signal peptide (domain A), as well as the N-terminal LRR flanking domain B and first three LRRs of the LRR domain C are conserved across all homologs (**Figure 6b**). Likewise, the C-terminal last 13 LRRs of domain C, as well as the acidic juxtamembrane domain D, the single-pass transmembrane domain E, and the intracellular domain F are highly conserved (**Figure 6b**).

The number of LRRs within the variable region of domain C vary widely between species and between homologs within species (**Figure 6b**). This intragenic recombination is the source of large variations in the number of LRRs in the *Hcr2* (for homologous to *Cladosporium* resistance gene *Cf-2*) gene family to which *Cf-2* belongs (Caicedo and Schaal, 2004). We identified five subclades of the *Hcr2* family. Two of these are the previously described *Hcr2* homologues evolving separately on chromosome 6, represented by *Cf-2* (*Hcr2* clade 5) and *Hcr2-2A* (*Hcr2* clade 4) (Dixon *et al.*, 1998) (**Figure 6a**). The clade to which *Cf-2* and *Cf-5* belongs only contains genes from species belonging to the *Solanum* subgenus *Solanum sensu stricto* section *Lycopersicoides* and *Lycopersicon*, indicating an evolutionary novel origin for this subclade. By contrast, the clade to which *Hcr2-2A* belongs additionally contains genes from potato (*Solanum* subgenus *Solanum sensu stricto* section *Petota*) (**Figure 6a**). These

genes all reside on the same locus on chromosome 6. There is one eggplant (*Solanum* subgenus *Leptostemonum* section *Melongena*) *Hcr2* gene which resides on chromosome 6, and which clusters with other eggplant *Hcr2* genes in a separate subclade (*Hcr2* clade 3) (**Figure 6a**). Finally, we identified two *Hcr2* clades which reside on chromosome 12 in tomato. *Hcr2* clade 2 is present is tomato and pepper species, while clade 1 is present in *Nicotiana* species as well (**Figure 6a**). The fact that clade 1 *Hcr2* homologs are widely conserved in the Solanaceae family indicates that this may represent the ancestral homolog (**Figure 6a**). Finally, we also identified a gene family present on chromosome 12 in tomato which contains homology to the 3' region of *Cf-2* which encodes for the conserved part of domain C, and domain D, E, and F, but with no homology in the remainder of the gene. This indicates that the signalling domain of the *Hcr2* gene family has been repurposed for this gene family.

Hcr2 homologs evolve by intragenic recombination at these distinct loci, giving rise to genes encoding for varying numbers of 24 amino acid LRRs (**Figure 6b**). The N-terminal and C-terminal encoding regions are highly conserved, but it is not excluded that sequence exchange occurs in the variable LRR region between *Hcr2* homologs on different genomic loci. To investigate whether the pattern of evolution seen for the *Hcr2* gene family is unique we looked at the *Hcr9* gene family in Solanaceae. The Hcr9 locus is the only other locus from which *R* genes against *C. fulvum* have been cloned. Interestingly, unlike *Hcr2* homologs *Hcr9* family members do not have the same repetitive gene structure on either a nucleotide or amino acid level (Dixon *et al.,* 1996). The *Hcr9* gene family is widely conserved in the Solanaceae family (**Figure S6a**). The evolutionary pattern is different from the *Hcr2* gene family: while the *Hcr2* gene family evolves by intragenic recombination in the variable LRR-encoding domain, giving rise to highly repetitive regions encoding for differing numbers of LRRs, the *Hcr9* family evolves by diversifying selection within the LRR encoding region. Unlike the *Hcr2* gene family, the *Hcr9* family aligns over the entire sequence, and most family members encode for

the same number of imperfect LRRs which do not display the same repetitiveness as the *Hcr2* family (Dixon *et al.*, 1996). Missing LRRs in some family members can be explained by large deletions (**Figure S6b**).

Cf-2 and *Cf-5* are the only identified *Hcr2* homologs with a known molecular function. The specificity determining region distinguishing *Cf-2* and *Cf-5* is the variable region of domain C (**Figure 6b**) (Seear and Dixon, 2003). *Cf-5* does not require *Rcr3* for it function (Dixon *et al.*, 2000), nor are the related PLCP genes *Pip1* or *Pfp1* required for triggering HR when Cf-5 and Avr5 are co-expressed by agroinfiltration in *N. benthamiana* (**Figure 1g**). Therefore, it appears that the evolutionary novel *Cf-2* is unique in its requirement for the ancient *Rcr3*.

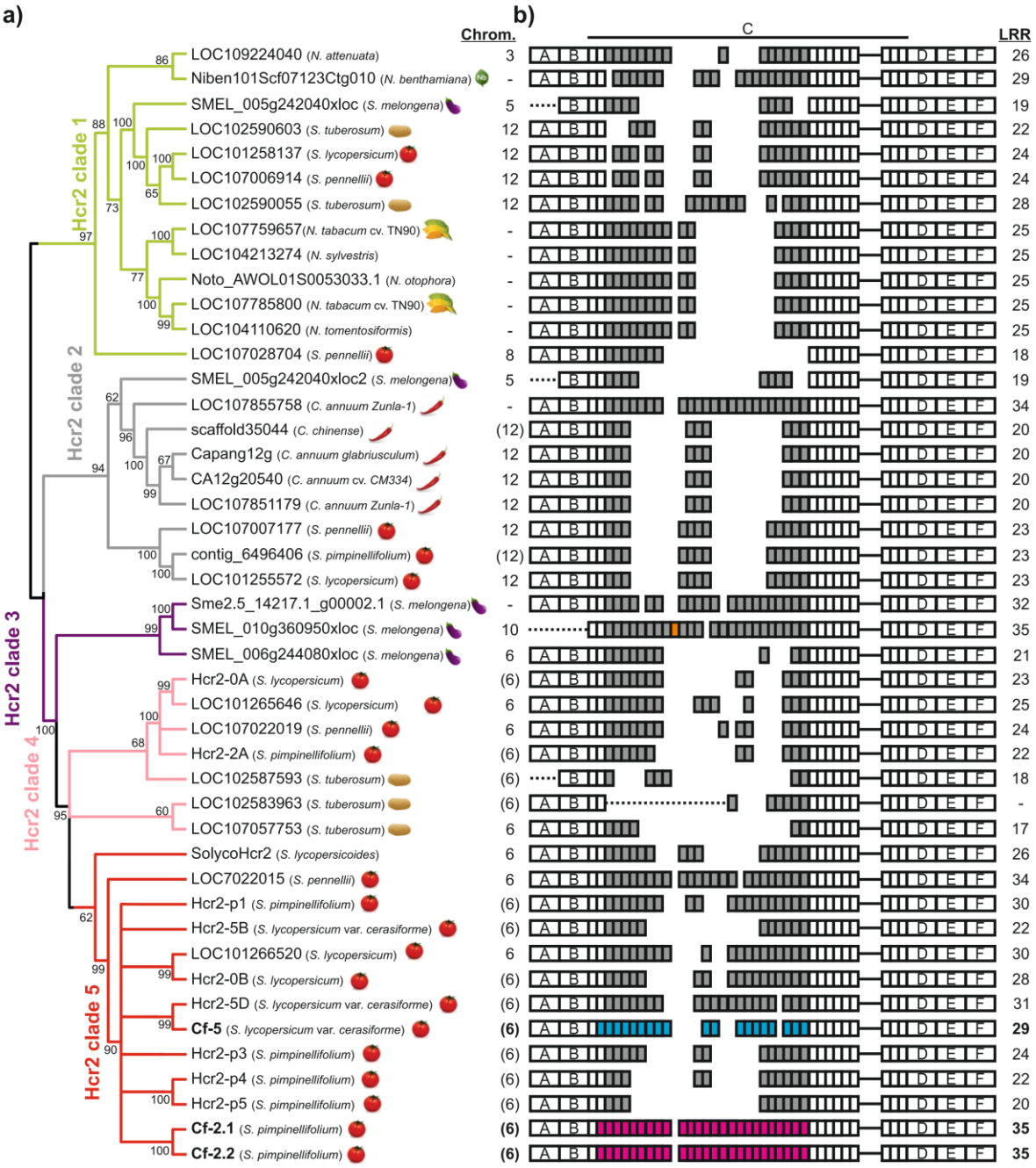


Figure 6: *Cf-2* evolved in a tomato-specific subclade of the *Hcr2* gene family. **a)** Phylogenetic tree of the *Hcr2* gene family based on the conserved N-terminal and C-terminal region (corresponding to amino acid residues 1 - 138 and 713 – 1112 from Cf-2). The evolutionary history was inferred using the Maximum Likelihood method based on the based on the Whelan and Goldman model. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. Shown are percentage bootstrap values. *Hcr2* family members with known recognition specificity are indicated in bold, the chromosome on which these genes reside is indicated, expected chromosome numbers are indicated in parentheses. **b)** Amino acid alignment comparison to Cf-2 using MUSCLE. LRRs of domain C were predicted using LRRfinder (Offord *et al.*, 2010). Numbers indicate total number of LRRs in domain C, grey bars indicate LRR region evolving by intragenic recombination, orange bars indicate additional LRRs not found in Cf-2. Magenta and cyan bars indicate LRR region in Cf-2 or Cf-5 which can confer Avr2 or Avr5 specificity in a Cf-5 or

Figure 6. (continued). Cf-2 background respectively. Dotted line indicates unknown sequence. A, signal peptide, B, N-terminal LRR flanking domain, C, LRR region, D, juxtamembrane region, E, transmembrane domain, F, cytoplasmic tail.

Discussion:

Here we show that *Cf-2* evolved in a *Solanum*-specific clade of the *Hcr2* family co-opting the pre-existing Rcr3 for Avr2 recognition. *Rcr3* is conserved in the Solanaceae, but only Rcr3 homologs from potato and tomato can effectively participate in Cf-2-dependent Avr2 recognition. This is despite these Rcr3 homologs interacting with Avr2, which indicates that the interaction of Rcr3 with Avr2 and Cf-2 are genetically separable. We found that residue Q284 in the *S. pimpinellifolium* Rcr3 homolog which co-evolved with Cf-2 is not the optimal residue for interaction with Avr2. Within tomato species Q284 occurs only in *S. pimpinellifolium* (Hörger *et al.*, 2012), and all *S. pimpinellifolium* lines containing *Cf-2* contain an Rcr3 variant with Q284 (S. malik, personal communication). It is tempting to speculate that this residue can interact with a tomato-derived factor, thereby triggering *Cf-2*-dependent autonecrosis.

Avr2 inhibition of Rcr3 involves at least three residues. Some of these residues may also be under selection due to interaction with other pathogen inhibitors. For example, the H148D substitution in $Rcr3^{lyc}$ is essential for inhibition by the *Phytophthora mirabilis* effector PmEPIC1 (Dong *et al.*, 2014). Combining our data with previously published data on Avr2 inhibition of PLCPs (van Esse *et al.*, 2008; Shabab *et al.*, 2008; Kaschani *et al.*, 2010) we can predict whether a PLCP is likely to be inhibited by Avr2 based on the following rules: 1) non-charged small residue at position 151, 2) asparagine at position 194, 3) positively charged residue at position 284 (**Figure S1, Figure S5, Table S3**). Additionally, since the interaction of Avr2 with Rcr3 occurs at acidic pH (< pH6.0), these PLCPs should be active at a low pH (Rooney *et al.*, 2005; Kaschani *et al.*, 2010).

Surprisingly we did not detect any HR upon the transient co-expression of Cf-2, Rcr3, and GrVAP1 in *N. benthamiana*. Whether this is due to differences in the experimental setup is not known. We know there was a genetic contamination in the tomato *Cf-2* line used in those experiments (Lozano-Torres *et al.*, 2012), which contained the autonecrosis-inducing $Rcr3^{lyc}$ allele. This could explain why that *Cf-2* line appeared more resistant to *G. rostochiensis*.

Cf-5-dependent recognition of Avr5 does not require Rcr3, or the closely related PLCPs Pip1 and Pfp1. Since the specificity-determining component of Cf-5 and Cf-2 lies in the variable LRR region we conclude that it is likely this region which interacts with Rcr3, and that Cf-2 is unique within the *Hcr2* family in its requirement for Rcr3. Therefore, we propose that when *Cf-2* evolved, Rcr3 gained a novel molecular function as a co-receptor for the recognition of Avr2. This contrasts with for example Prf and ZAR1: two highly conserved NLRs in Solanaceae and eudicots, respectively. Prf recognizes the interaction of the *P. syringae* effectors AvrPto and AvrPtoB with the tomato kinase Pto (Ntoukakis *et al.*, 2014). Since *Prf* is more widely conserved than *Pto* (Bombarely *et al.*, 2012) it is interesting to speculate that *Pto* evolved to interact with Prf for the recognition of different pathogen effectors. While transgenic *N. benthamiana* lines carrying *Pto* driven by the 35S promoter are resistant to *P. syringae* pv. *tabaci* strains expressing *AvrPto* (Rommens *et al.*, 1995) and the development of a HR in these lines depends on the endogenous *N. benthamiana Prf* (Lu *et al.*, 2003), this line does not confer resistance to *P. syringae* expressing *AvrPtoB* (Balmuth and Rathjen, 2007). Additionally, expression of *Pto* from its native promoter in transgenic *N. benthamiana* lines does not lead to resistance to *P. syringae* strains carrying either *AvrPto* or *AvrPtoB*, but *N. benthamiana* transgenics lines expressing both *Pto* and tomato *Prf* from their endogenous promoters are resistant against these *P. syringae* strains (Balmuth and Rathjen, 2007). This leads to the conclusion that *Pto* has evolved to interact with the *Prf* homolog present in tomato, but that overexpression of *Pto* can force the interaction with other *Prf* homologs for the

recognition of AvrPto, but not AvrPtoB. Like Prf, Arabidopsis ZAR1 is a NLR that interacts with the receptor-like cytoplasmic kinases (RLCKs) ZED1 (Lewis *et al.*, 2010; Lewis *et al.*, 2013), ZRK3 (Seto *et al.*, 2017), and RKS1 (G., Wang *et al.*, 2015) for the recognition of the *P. syringae* effectors HopZ1a and HopF2a, and the *Xanthomonas campestris* effector AvrAc, respectively. The *N. benthamiana* ZAR1 interacts with the RLCK JIM2 for the recognition of the *X. perforans* effector XopJ4 (Schultink *et al.*, 2018). Co-expression of ZED1 with HopZ1a by agroinfiltration is sufficient to trigger a HR in *N. benthamiana*, and this HR depends on the endogenous *N. benthamiana ZAR1* (Baudin *et al.*, 2017). Overexpression of the Arabidopsis ZAR1, however, with XopJ4 in a *N. benthamiana zar1* knock-out line is not sufficient to trigger a HR (Schultink *et al.*, 2018). The RLCKs *ZED1*, *RKS1*, *ZRK3*, and *JIM2* appear to be lineage-specific and have therefore likely evolved to interact with the available *ZAR1* allele.

In conclusion, here we highlight how the evolution of a new receptor can give an ancient gene a new function. This receptor has evolved very recently, unlike *Prf* and *ZAR1* which work in a similar way. It is not unlikely that these other receptors initially evolved in a similar way, and subsequently their recognition spectra expanded by the evolution of additional decoys. Understanding the evolution of defence components is crucial for the future rational engineering and application of these genes.

Material & Methods:

Cloning of general binary cloning vectors:

The Golden Gate Modular Cloning (MoClo) kit (Weber *et al.*, 2011) and the MoClo plant parts kit (Engler *et al.*, 2014) was used for cloning, and all vectors are from this kit unless specified otherwise. The general binary cloning vector pJK001c (pL1V2-F1) was generated by combining pICH47732, pICH41722, and pICSL4723 (Addgene #48000, #48016, #86172,

respectively) in a BpiI Golden Gate reaction. pJK263 (pL1M-R1-pMAS::P19) was created by combining pJK046 (pL0M-SC-P19) (Grosse-Holz *et al.*, 2018) with pICH41432 (OCS terminator) and pICH47802 (Addgene #50343 and #48007, respectively), and EC15317 (MAS promoter; ENSA consortium) in a BsaI Golden Gate reaction. pJK241 (pL1M-F1-FASTred) was constructed by combining pICH47732 (Addgene #48000) with pICSL70008 (Addgene #50336) in a BsaI Golden Gate reaction. The general binary cloning vector pJK268c (pL1V2-P19-F2) containing the P19 silencing inhibitor and the general binary cloning vector pJK242c (pL1V2-FAST-F2) containing the FAST selection module was constructed by combining either pJK263 (pL1M-R1-pMAS::P19) or pJK241 (pL1M-F1-FASTred) respectively with pICH47742, pICH41744, and pICSL4723 (Addgene #48001, #48017, and #86172, respectively) in a BpiI Golden Gate reaction. The general binary cloning vector pJK497 (pL0V2-P19-2x35S::PR1aSP-C) containing the P19 silencing inhibitor and allowing for cloning of a single gene with stop-codon and without N-terminal signal peptide in between a 2x35S promoter, NtPR1a signal peptide with intron, and 35S terminator was constructed by combining pICH41264 (Addgene #47993), pJK002 (pL0M-S-NtPR1a-PIV2 (Grosse-Holz *et al.*, 2018), pICH51288 and pICH41414 (2x35S and 35S terminator; Addgene #50269 and #50337), and pJK268c (pL1V2-P19-F2) in a BsaI Golden Gate reaction.

Cloning of general level 0 modules

pJK369 (pL0M-C-Cf2), pJK541 (pL0M-C-Cf5), pJK515 (pL0M-C-Hcr2-2A), pJK522 (pL0M-C-Avr5), pJK401 (pL0M-C-NbRcr3$_{CF}$) were synthesized as GeneArt Genestring (Thermofisher, **Table S5**) without the nucleotides encoding N-terminal signal peptides and combined with pICH41264 (Addgene #47993) in a BpiI Golden Gate reaction. pJK369 (pL0M-C-Cf2), pJK541 (pL0M-C-Cf5), pJK515 (pL0M-C-Hcr2-2A) were codon-optimized. pJP003 (pL0M-C-Pip1), pJP004 (pL0M-C-Rcr3pim), and pJP005 (pL0M-C-Rcr3lyc) were made by PCR and with pICH41264 (Addgene #47993) in a BpiI Golden Gate reaction. pJK019 (pL0M-

C-Avr2) was generated by cloning of Avr2 (Genbank accession AJ421629.1) lacking the first 20 amino acids constituting the N-terminal signal peptide from plasmid pFK3 (Shabab *et al.*, 2008) using the forward 5'-ttgaagacaaaggtgccaaaaaactacctggctgc'3' and reverse 5'-ttgaagacaaaagctcaaccgcaaagaccaaaacagc-3' primers into pICH41264 (Addgene #47993) in a BpiI Golden Gate reaction. pSM001 (pL0M-C-Avr2Δ6) was generated by PCR on pJK019 as a template using the forward 5'-ttgaagacaatgctgagctttgagaccacgaag-3' and reverse 5'-ttgaagacaaagcaaaggtgattaccctgcac-3' primers and was circularized in a BpiI Golden Gate reaction. pJK021 (pL0M-S-NtPR1a) was generated by combining an oligo-annealed primer pair (forward 5'-aatgggatttgttctcttttcacaattgccttcatttcttcttgtctctacacttctcttattcctagtaatatcccactcttgccgtgc-3' and reverse 5'-acctgcacggcaagagtgggatattactaggaataagagaagtgtagagacaagaagaaatgaaggcaattgtgaaaagagaacaaatcc-3') with pICH41258 (Addgene #47987) in a BpiI Golden Gate reaction.

Cloning of general level 2 binary vectors

pJK050c (pL2M-2x35S::P19) was generated by combining pJK046 (Grosse-Holz *et al.*, 2018) with pICH51288 and pICH41414 (2x35S and 35S terminator; Addgene #50269 and #50337, (Engler *et al.*, 2014)), and pJK001c (pL1V2-F1) in a BsaI Golden Gate reaction. pJK404 (pL2M-FAST-2x35S::Cf-2) was created by combining pJK369, pJK002 (Grosse-Holz *et al.*, 2018), pICH51288 and pICH41414 (2x35S promoter and 35S terminator; Addgene #50269, #50337), and pJK242c in a BsaI Golden Gate reaction. pJK024 (pL2M-P19-2x35S::Cf-2), pJK225 (pL2M-P19-2x35S::Avr2), pJK325 (pL2M-P19-2x35S::Avr2Δ6), pJK284 (pL2M-P19-2x35S::Rcr3pim), pJK509 (pL2M-P19-2x35S::Rcr3lyc), pJK488 (pL2M-P19-2x35S::Pip1), pJK490 (pL2M-P19-2x35S::Rcr3pim_C24AC25A) were generated by combining either pJK369 (pL0M-C-Cf2_GeneArt), pJK019 (pL0M-C-Avr2), pSM001 (pL0M-C-Avr2Δ6), pJP004 (pL0M-C-Rcr3pim), pJP005 (pL0M-C-Rcr3lyc), pJP003 (pL0M-

C-Pip1), or pJK150 (pL0M-C-Rcr3pimC24AC25A) with pJK002 (Grosse-Holz *et al.*, 2018), pICH51288 and pICH41414 (2x35S and 35S terminator; Addgene #50269 and #50337) and pJK268c (pL1V2-P19-F2) in a BsaI Golden Gate reaction.

Sequence retrieval

Tomato *Rcr3* (*Solyc02g076980.2*), *Pip1* (*Solyc02g077040.2*), *PIP1-like* (*Solyc03g006210.1*), *Pfp1* (*Solyc02g077050.2*), *Cf-9* (Genbank LPU15936) and the 3' conserved region of *Cf-2* (Genbank U42445.1) were used to retrieve homologous genes in other organisms by BLAST searches in Solgenomics against the available genomes and databases. Additional sequences were obtained by BLAST searches against NCBI. Sequences which had not been annotated or were incorrectly annotated were manually annotated. All sequences are described in **Table S1**. *N. benthamiana* PLCP genes were manually re-annotated and compared to those found in the recently re-annotated *N. benthamiana* gene-models (Kourelis *et al.*, 2018).

Tree building

Clustal Omega (Sievers *et al.*, 2011; Li *et al.*, 2015) was used to align sequences. Determining the best model for maximum likelihood phylogenetic analysis and the phylogenetic analysis was performed in MEGA X as indicated in the text (Kumar *et al.*, 2018).

E. coli expression vector cloning and Avr2 purification

The partially domesticated pET28b vector pJK082 (pL1V-pET28b), compatible with level 1 Golden Gate assembly of transcription units using BsaI, was generated by combining a PCR fragment derived from pET28b (EMD Biosciences) using the forward 5'-ttggtctcactccatttcgcgggatcgagatctcg-3' and reverse 5'-ttggtctcacgcttggcgaatgggacgcgccctg-3' primers and a PCR fragment derived from pFK210B-AtMIR390a-B/c (Addgene #51777 (Carbonell *et al.*, 2014)) using the forward 5'-aaagcgtgagaccattaggcaccccagg-3' and reverse 5'-aaggagtgagaccgtcgaggtgcagactgg-3' primers in a BsaI Golden Gate reaction. The N-

terminal tagging module for addition of an OmpA signal peptide fused to a 6xHis tag followed by a TEV protease cleavage site pJK120 (pL0M-S-OmpA-His6-TEV) was generated by combining a oligo-annealed primer pair (forward 5'-aatgaaaaagacagctatcgcgattgcagtggcactggctggtttcgctaccgttgcgcaagctcatcaccatcaccatcacgaaaacctatacttcca-3' and reverse 5'-accttggaagtataggttttcgtgatggtgatggtgatgagcttgcgcaacggtagcgaaaccagccagtgccactgcaatcgcgatagctgtcttttt-3') with pICH41258 (Addgene #47987) in a BpiI Golden Gate reaction. The T7 promoter module pJP001 (pL0M-PU-T7-LacO-T7) was generated by combining a oligo-annealed primer pair (forward 5'-ggagtaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatatac-3' and reverse 5'-cattgtatatctccttcttaaagttaaacaaaattatttctagaggggaattgttatccgctcacaattcccctatagtgagtcgtatta-3') with pICH41295 (Addgene #47997) in a BpiI Golden Gate reaction. The T7 terminator module pJP002 (pL0M-T-T7 terminator) was generated by combining a PCR fragment derived from pET28b (EMD Biosciences) using the forward 5'-ttgaagacaagcttggatccggctgctaacaaagc-3' and revers 5'-ttgaagacaaagcgatccggatatagttcctcctttc-3' primers with pICH41276 (Addgene #47994) in a BpiI Golden Gate reaction. The expression vector for periplasmic secretion of 6xHis-tagged Avr2 (pJK153, pET28b-T7::OmpA-HIS-TEV-Avr2), or Avr2Δ6 (pSM153, pET28b-T7::OmpA-HIS-TEV-Avr2Δ6), were generated by combining pJK019 (pL0M-C-Avr2) or pSM001 (pL0M-C-Avr2Δ6) with pJK082 (pL1V-pET28b), pJK120 (pL0M-S-OmpA-His6-TEV), pJP001 (pL0M-PU-T7-LacO-T7), and pJP002 (pL0M-T-T7 terminator) in a BsaI Golden Gate reaction. For protein expression pJK153 (pET28b-T7::OmpA-HIS-TEV-Avr2) and pSM153 (pET28b-T7::OmpA-HIS-TEV-Avr2Δ6) were transformed to *E. coli* strain Rossetta2(DE3)pLysS (Novagen/Merck). An overnight grown starter culture was diluted 1/100 in TB [per liter: 24 gr yeast extract, 20 gr Bacto-tryptone, 4 ml glycerol, after autoclaving add

0.017 M KH_2PO_4 and 0.072 M K_2HPO_4 (final concentrations)], grown for approximately 3 hrs to $OD_{600} \approx 0.6$ and induced overnight at 37°C by the addition of 0.8 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Afterwards the supernatant was collected by centrifugation of the bacterial culture at 10,000 x g, 4°C for 20 min. The supernatant was adjusted to 150 mM NaCl, 30 mM imidazole and run over a Ni- NTA agarose (Qiagen)-loaded column by gravity-flow. The Ni-NTA agarose was washed with 10 column volumes of washing buffer (50 mM Tris-HCl, 150 mM NaCl, 50 mM imidazol, pH 8) and the purified protein was eluted in elution buffer (50 mM Tris-HCl, 150 mM NaCl, 250 mM imidazol, pH 8). Finally, the protein was concentrated by centrifugation in 3kDa cut-off filters (Amicon Ultra-15 Centrifugal Filter Units, Merck) at 4,500x g, 4°C, flash-frozen in liquid nitrogen and stored at -80°C. Protein concentration was determined by Bradford assay using BSA as a standard, and purity was verified by running the protein on 18% SDS-PAGE gels followed by Coomassie staining (Pink *et al.*, 2010).

Agroinfiltration

For transient expression of proteins in *N. benthamiana*, *A. tumefaciens* strain GV3101 pMP90 carrying binary vectors were inoculated from glycerol stock in LB supplemented with 25 µg/ml rifampicin, 50 µg/ml gentamycin and 50 µg/ml kanamycin and grown O/N at 28°C until saturation. Cells were harvested by centrifugation at 2000x g, RT for 5 min. Cells were resuspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES-KOH pH 5.6, 200 µM acetosyringone) to an OD_{600} of 0.25 unless stated otherwise and left to incubate in the dark for 2 hours at RT prior to infiltration into 5-week-old *N. benthamiana* leaves.

HR assays

For HR assays 5-week-old *N. benthamiana* plants were agro-infiltrated with the stated combinations of binary vector containing Agrobacterium at $OD_{600} = 0.25$ each. Two sites were

infiltrated as described above, making sure to always include a positive control in each leaf. Images were taken 5 dpi unless stated otherwise.

ABPP

Apoplastic fluid (AF) was isolated from agroinfiltrated *N. benthamiana* plants expressing different proteins at 2-3 days post-infiltration according to (Grosse-Holz *et al.*, 2018). Briefly, the AF was extracted by vacuum infiltrating *N. benthamiana* leaves with ice-cold MilliQ. Leaves were dried with filter paper to remove excess liquid, and apoplastic fluid was extracted by centrifugation of the leaves in a 20 ml syringe barrel (without needle or plunger) in a 50 ml falcon tube at 2000x g, 4°C for 25 min. Samples were used directly in labelling reactions. Samples were adjusted to 5 mM DTT, 50 mM NaAc pH 5, and pre-inhibited for 45 min with either 100 µM E64 (a general PLCP inhibitor), 2-4 µM purified Avr2 or Avr2Δ6, or 1% DMSO (final concentration) as a control. After pre-inhibition, samples were labelled with 2 µM MV201 (a fluorescent E64 derivative (Richau *et al.*, 2012)) for 3 hours at room temperature (total volume 50 µl). The labelling reaction was stopped by precipitation with 5 volumes of ice-cold acetone, followed by a 10 s vortex and immediate centrifugation at 16,100x g, 4°C for 5 min. Samples were resuspended in 2x sample buffer [100 mM Tris-HCl (pH 6.8), 200 mM DTT, 4% SDS, 0.02% bromophenol blue, 20% glycerol] and boiled for 5 min 95°C prior to running on 15% SDS-PAGE gels. Fluorescence scanning was performed on an Amersham Typhoon scanner (GE Healthcare), using Cy3 settings. Equal loading was verified by Coomassie staining (Pink *et al.*, 2010). Quantification of ABPP gels was done in ImageJ.

CRISPR/Cas9 construct design

Two CRISPR/Cas9 sgRNAs targeting *NbPip1* (*Niben101Scf08921g02024.1*) at two positions 90bp apart were designed to minimize *in silico* off-targets (Stemmer *et al.*, 2015) with maximum *in silico* on-target activity (Doench *et al.*, 2016). Specific PCR products were

generated by amplifying from the template for sgRNAs pICH86966 (Addgene #46966 (Nekrasov *et al.*, 2013)) using forward primers 5'-tgtggtctcaattgtgactgcatgattcagcgtgttttagagctagaaatagcaag-3' (for sgRNA_25) or 5'-tgtggtctcaattgaaaattgttagagatgccggttttagagctagaaatagcaag-3' (for sgRNA_44) and reverse primer 5'-tgtggtctcaagcgtaatgccaactttgtac-3'. The two PCR products were combined with the AtU6 promoter (pICSL01009, Addgene #46968, (Nekrasov *et al.*, 2013)), and pICH47751 or pICH47761 (Addgene #48002 and #48003) in a Golden Gate reaction with BsaI to generate pJK084 (pL1M-F3-AtU6p::sgRNA>NbPip1_25) and pJK085 (pL1M-F3-AtU6p::sgRNA>NbPip1_44) respectively. The final binary CRISPR/Cas9 vector (pL2M-KAN-Cas9>NbPip1) was constructed by combining pAGM4723 and pICH41780 (Addgene #48015 and #48019) with pICSL11017 (Addgene #51144), pICH47742 (Addgene #49771, (Belhaj *et al.*, 2013)), and pJK084 and pJK085 in a Golden Gate reaction with BpiI, resulting in pJK090.

sgRNA_25 (NbPip1 position 1) underlined: spacer sequence

gtgactgcatgattcagcgtgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagt
cggtgctttttt

sgRNA_44(NbPip1 position 2) underlined: spacer sequence

gaaaattgttagagatgccggttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagt
cggtgctttttt

Nicotiana benthamiana tissue culture

Agrobacterium mediated transformation of *N. benthamiana* was performed as described (Sparkes *et al.*, 2006). Briefly, 4-week old *N. benthamiana* plants were agroinfiltrated with *A. tumefaciens* strain GV3101 pMP90 carrying binary plasmid pJK090 (pL2M-KAN-Cas9>NbPip1) at OD_{600} = 0.2. Three days post-infiltration, infiltrated leaves were cut into 1

cm^2 sections and were cultured under sterile conditions on shooting media [per liter: 4.3 gr MS basal salts (1x), 30 gr sucrose, 1x B5 vitamins, 0.1 mg Naphthalene Acetic Acid (NAA), 0.59 gr MES, pH=5.7 using KOH, 4 gr Agargel, after autoclaving add 1 mg Benzylaminopurine (BAP), 300 mg Ticarcillin/clavulanic acid] for three days, and subsequently transferred to shooting media supplemented with 150 μg/ml kanamycin to select for stable transformants and were further cultured until shoots appeared. Shoots >3 mm were excised and transferred to rooting media [per liter: 2.15 gr MS basal salts (0.5x), 5 gr sucrose, 0.1 mg NAA, pH=5.8 with KOH, 2.5 gr Gelrite, after autoclaving add 300 mg Ticarcillin/clavulanic acid] supplemented with 150 μg/ml kanamycin and were further cultured until roots appeared, after which plantlets were transferred to soil and grown in the growth chamber. T0 *N. benthamiana* transformants were screened for *NbPip1* mutant alleles by sanger sequencing of PCR fragments generated using forward primer: 5'- actaatctcactttcggagcaag-3' and reverse primer 5'-ccacaattcttgtcgaacaacct-3'. *NbPip1* mutant T1 and T2 lines lacking the T-DNA containing the kanamycin resistance and *Cas9* transgenes were confirmed by PCR on the *NPTII* using forward primer: 5'-gaggctattcggctatgactgg-3' and reverse primer 5'-tatgtcctgatagcggtccg-3' and these lines were used for further experiments.

Virus-induced Gene Silencing (VIGS) cloning and silencing

A partially domesticated TRV2 vector compatible with Golden Gate cloning using BsaI (pJK037) was generated by combining the PCR fragments derived from the plasmid template pTRV2 (Liu *et al.*, 2002) using the forward 5'-ttggtctcaggatccggtaccgagctcacgc-3' and reverse 5'-ttggtctcaaggccttttcgacctttttcccctgc-3' and forward 5'-ttggtctcagcctctttcctgtggatagcacgtac-3' and reverse 5'-ttggtctcagaattcggtaaccttactcacagaatc-3' primers with the PCR fragment derived from the plasmid template pFK210B-AtMIR390a-B/c (Addgene #51777 (Carbonell *et al.*, 2014)) using the forward 5'-aaaaagaattcagagaccattaggcaccccagg-3' and reverse 5'-aaaaaggatcctgagaccgtcgaggtgcagactgg-3' in a BsaI Golden Gate reaction. Two independent

VIGS constructs targeting *NbPip1* were ordered as GeneArt GeneStrings (ThermoFisher) and combined with pJK037 in a BsaI Golden Gate reaction generating pJK220 (pL2M-TRV2>NbPip1#1) containing a 300bp fragment targeting *NbPip1* and pJK221 (pL2M-TRV2>NibenPip1#2) containing a previously published 576bp fragment targeting *NbPip1* (Xu *et al.*, 2012). *A. tumefaciens* strain GV3101 pMP90 carrying binary plasmid pJK220 (pL2M-TRV2::NbPip1#1), pJK221 (pL2M-TRV2::NbPip1#2), or *TRV2::GFP* were mixed 1:1 with bacteria containing binary plasmid TRV1 (Liu *et al.*, 2002) at a combined final OD_{600} = 1 in infiltration buffer as described above. The mixed cultures were infiltrated into leaves of 10-day-old *N. benthamiana* plants. The infiltrated seedlings were grown for 3-4 weeks in a growth chamber until use.

Colony count

Pseudomonas syringae pv. *tomato* DC3000 ΔhopQ1-1 (Wei *et al.*, 2007) was grown to saturation from glycerol stock O/N at 28°C in LB supplemented with 25 µg/ml rifampicin, washed and resuspended in 10 mM $MgCl_2$ at OD_{600} = 0.0002 (1 x 10^5 CFU/ml) and were used for infection of plants by syringe infiltration. For CRISPR/Cas9 *NbPip1* KO lines, four-week-old mutant or wild-type plants were infected, whereas for VIGS silencing of *NbPip1* plants were infected three weeks post infection with either pJK220 (pL2M-TRV2::NbPip1#1), pJK221 (pL2M-TRV2::NbPip1#2), or *TRV2::GFP*. Three leaf discs (8 mm diameter) were punched from the infected leaves of four independent plants for each treatment at 3 hrs, 24 hrs, 48 hrs and 72 hrs post infection and were incubated in 15% H_2O_2 for 5 min to sterilise leaf surfaces. Leaf discs were washed twice in sterile MilliQ and ground in 1 ml sterile MilliQ for 5 minutes by addition of two metal beads in a tissuelyser (Biospec Products, Bartelsville, OK, USA). Serial dilutions of the homogenate were plated onto LB agar plates containing 25 µg/ml rifampicin. Colonies were counted after 48 h incubation at 28°C. Statistics were performed in R.

Acknowledgements.

We thank Matthieu Joosten for supplying the Cf-2::eGFP and Cf-4::eGFP plasmids; Ursula Pyzio for excellent plant care; John Baker for photography; Jonathan Jones, Sophien Kamoun, James Carrington, Sylvestre Marillonnet, and Nicola Patron for providing plasmids via AddGene.

Funding. This work has been supported by 'The Clarendon Fund' (JK), the Oxford Indira Gandhi Graduate Scholarship (SM), and the ERC Consolidator grant 616449 'GreenProteases' (RvdH, JK). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests. The authors have declared that no competing interests exist.

Author contributions. Conceptualization: JK, RvdH; Formal analysis: JK; Investigation: JK, SM; OM, SK, and PSK under supervision of JK; Funding acquisition: RvdH; Writing: JK, RvdH.

Supplemental figures

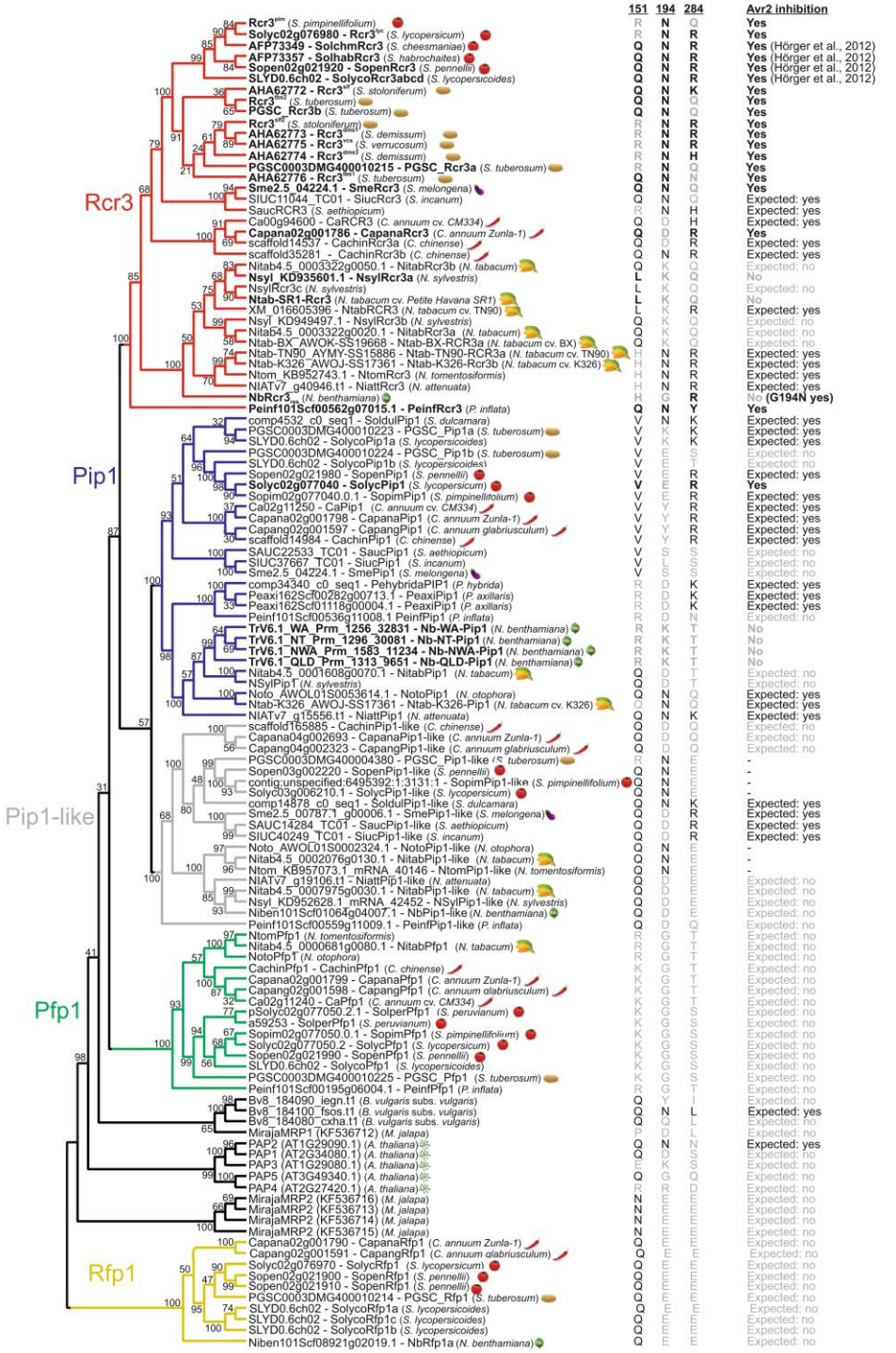


Figure S1: Tree of Pip1, Pfp1, Rcr3, Pip1-like PLCPs. The evolutionary history of the Rcr3 clade of proteases was inferred using the Maximum Likelihood method based on the Whelan and Goldman model. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. Shown are percentage bootstrap values. Genes which have been tested for Avr2 inhibition are indicated in bold. Amino acid residues corresponding to Rcr3^{lyc} R151, N194, and R284 are indicated. Residues expected to negatively influence Avr2 inhibition are indicated in grey.

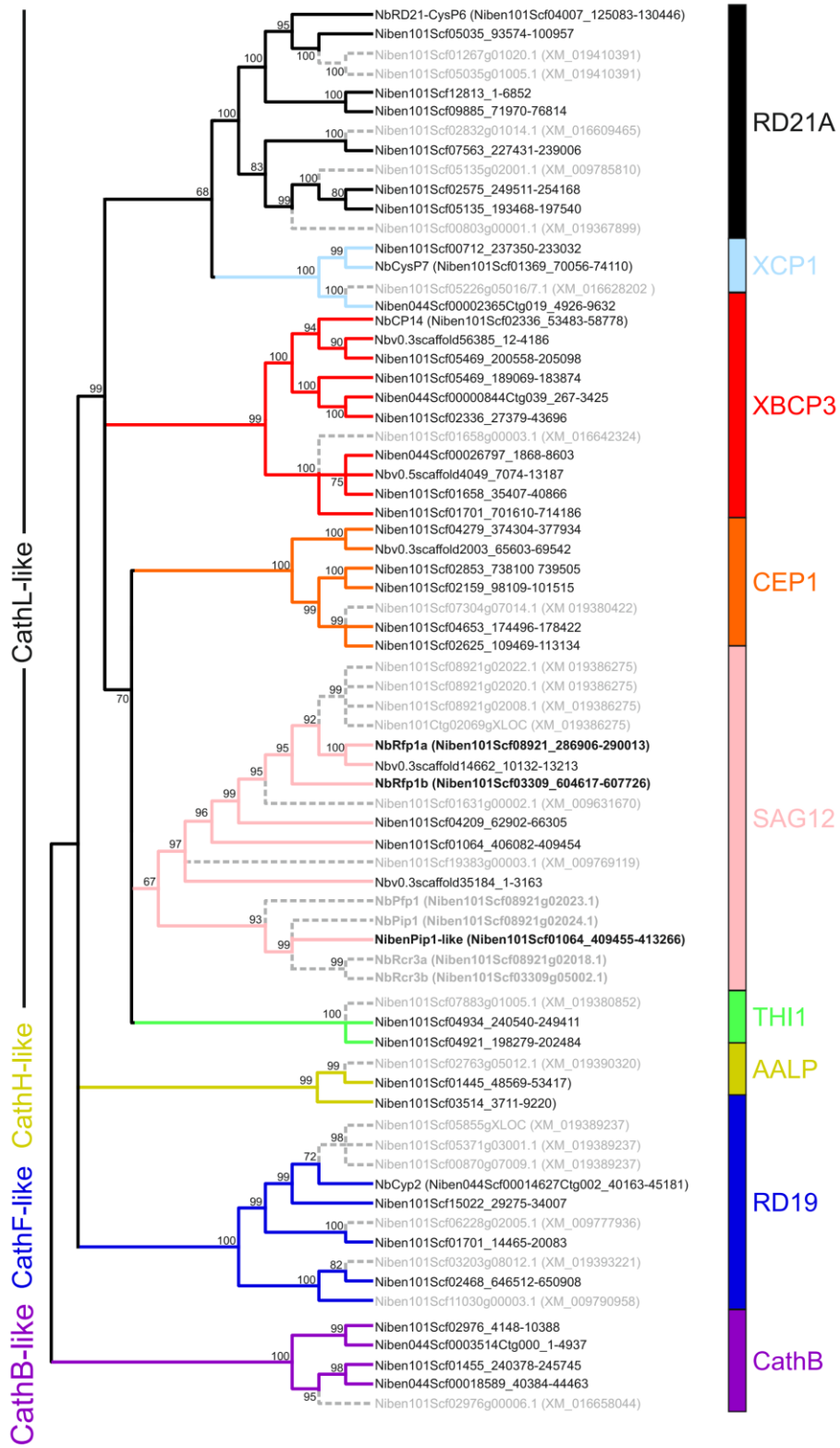


Figure S2: Tree of *N. benthamiana* PLCPs. The evolutionary history of the PLCP gene family was inferred using the Maximum Likelihood method based on the Whelan and Goldman model. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. Shown are percentage bootstrap values. Putative pseudogenes are indicated in grey. Naming of PLCP clades according to (Richau *et al.*, 2012).

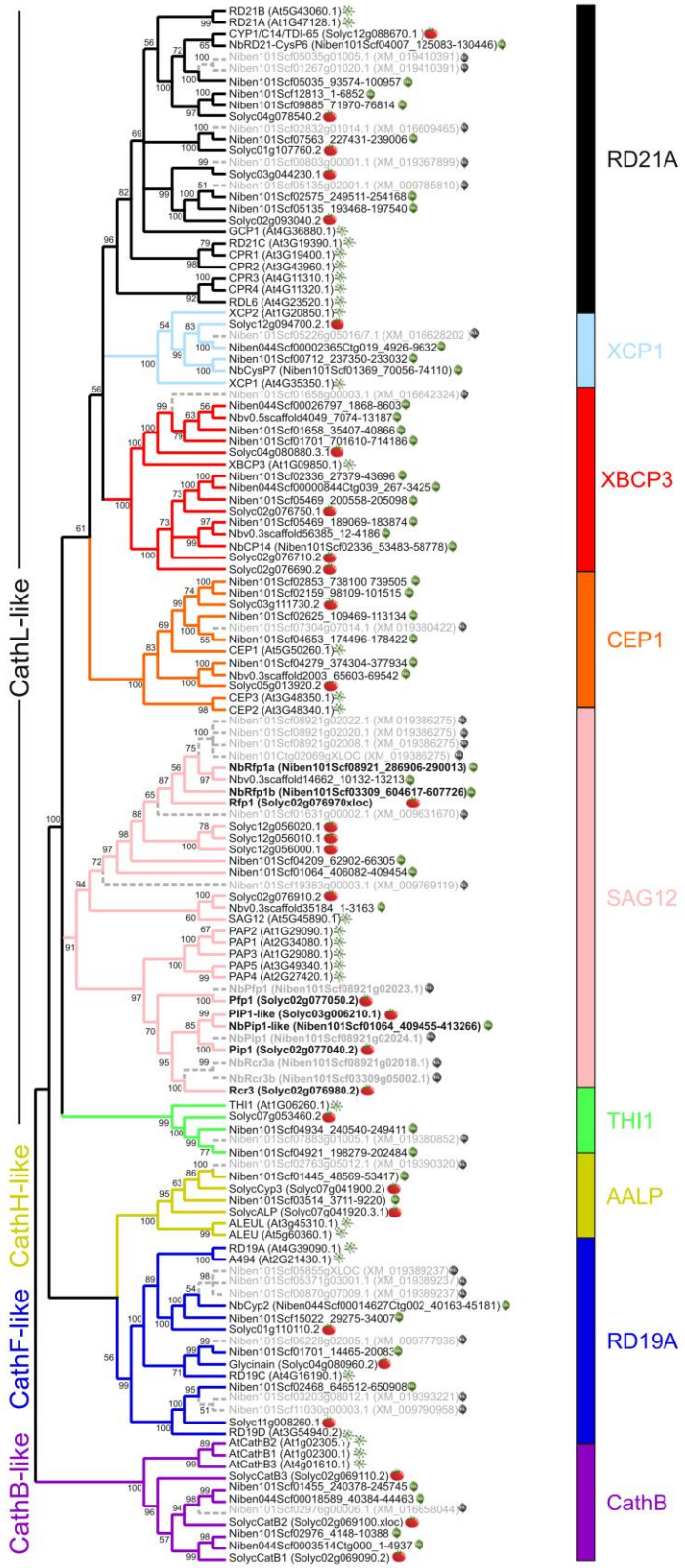


Figure S3: Tree of tomato, Arabidopsis, and *N. benthamiana* PLCPs. The evolutionary history of the PLCP gene family was inferred using the Maximum Likelihood method based on the Whelan and Goldman model. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. Shown are percentage bootstrap values. Putative pseudogenes are indicated in grey. Naming of PLCP clades according to (Richau *et al.*, 2012).

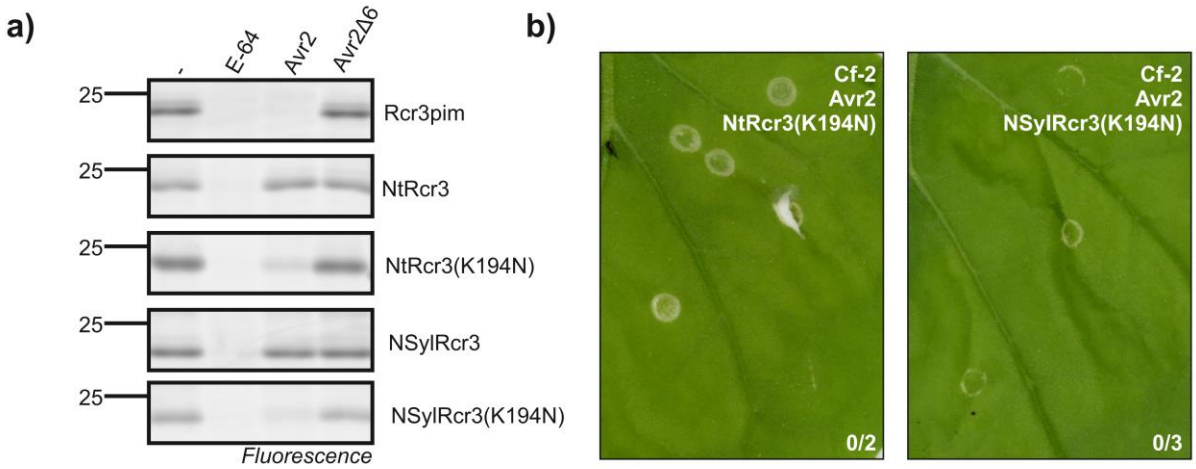


Figure S4: Rcr3 is an active protease in Solanaceae. **a)** Activity-based protein-profiling of Papain-like cysteine proteases. Rcr3^{pim}, and the *Nicotiana tabacum* and *Nicotiana sylvestris* Rcr3 homologs (NtRcr3 and NSylRcr3, respectively) and the K194N mutants thereof were co-expressed with the silencing inhibitor P19 in *N. benthamiana* leaves by agroinfiltration. Apoplastic fluids were isolated at 4 dpi, pre-incubated for 45 min with or without 100 µM E-64, 2 µM Avr2 or Avr2Δ6, or DMSO, and labelled for 3 hours with 2 µM MV201. Samples were separated on SDS-PAGE gels and scanned for fluorescence. **b)** Co-expression of Cf-2, Avr2 and the NtRcr3(K194N) or NSylRcr3(K194N) mutant protease in *N. benthamiana* by agroinfiltration. Binary vectors contained the silencing inhibitor P19 in the backbone, and the *A. tumefaciens* cultures were diluted to OD_{600} = 0.25 each in 1:1:1 mix. Pictures taken at 5 dpi. Numbers indicate the number of leaves showing HR.

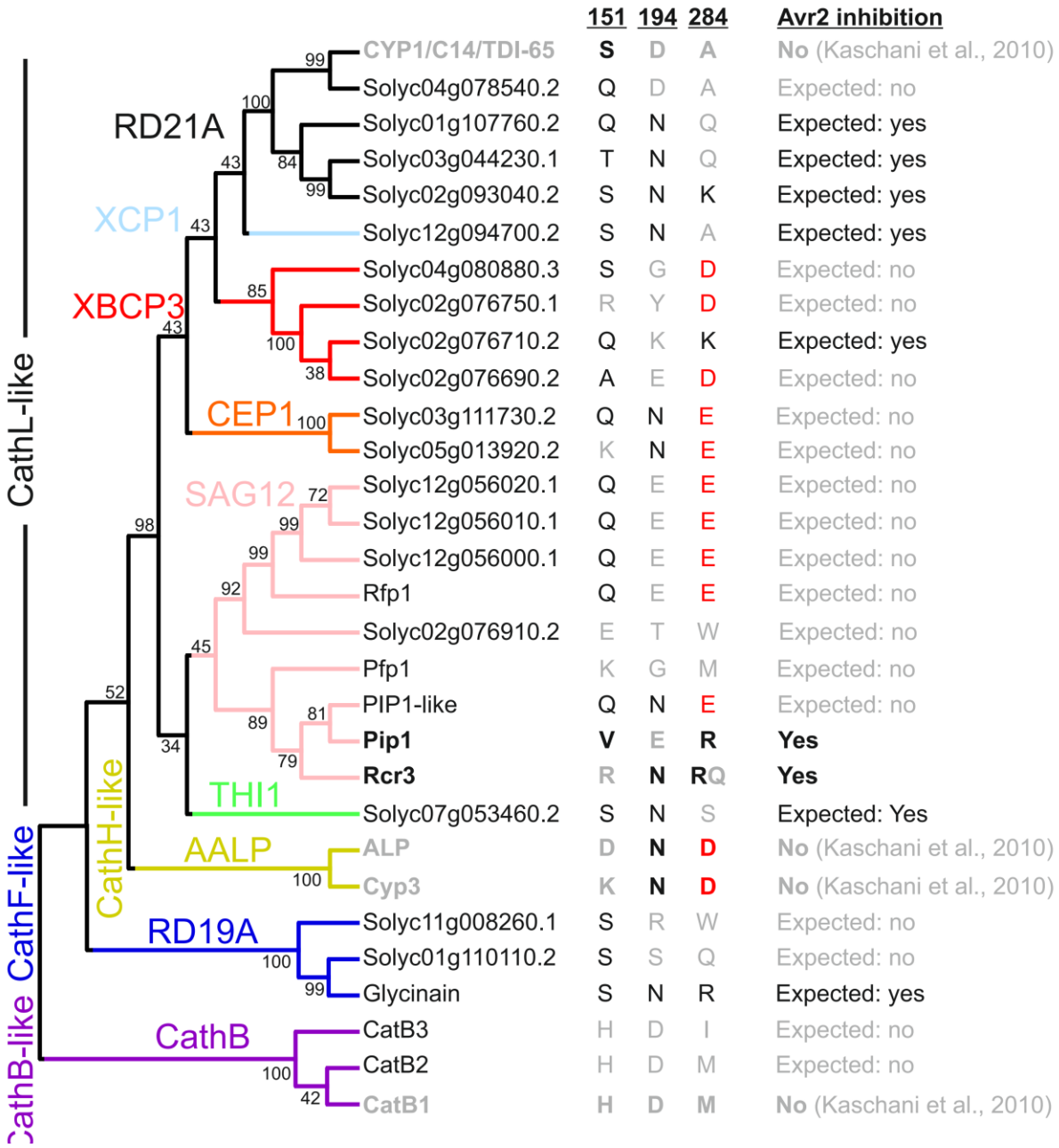


Figure S5: Tree of tomato PLCPs. The evolutionary history of the tomato PLCP clade of proteases was inferred using the Maximum Likelihood method based on the Whelan and Goldman model. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. Shown are percentage bootstrap values. Genes which have been tested for Avr2 inhibition are indicated in bold. Amino acid residues corresponding to Rcr3^{lyc} R151, N194, and R284 are indicated. Residues expected to negatively influence Avr2 inhibition are indicated in grey.

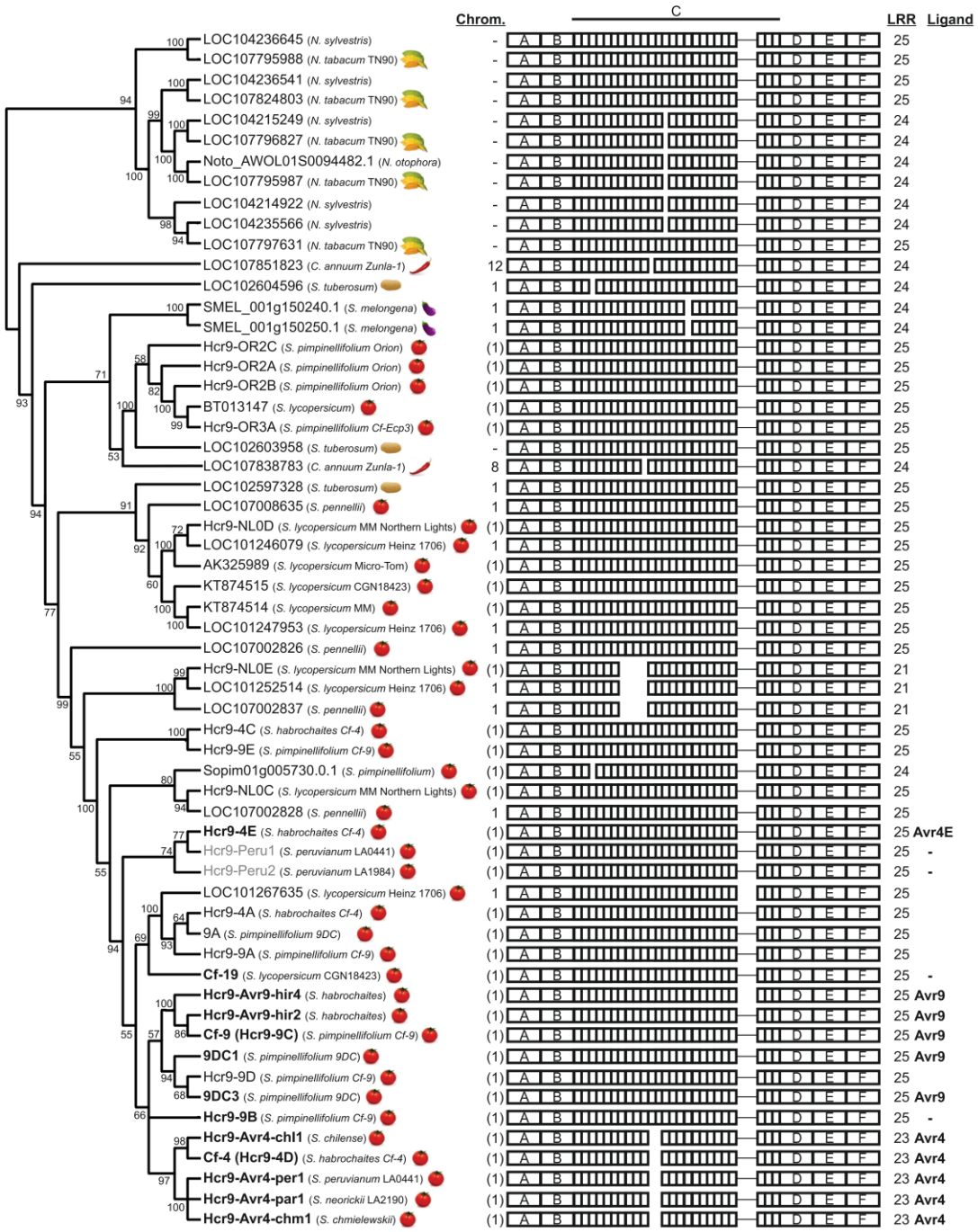


Figure S6: Hcr9 homologs are conserved in the Solanaceae. **a)** Phylogenetic tree of the *Hcr9* gene family. The evolutionary history was inferred by using the Maximum Likelihood method based on the based on the Whelan and Goldman model. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. Shown are percentage bootstrap values. Hcr9 family members with known recognition specificity are indicated in bold, in light grey those that induce autonecrosis upon agroinfiltration in *N. benthamiana*, the chromosome on which these genes reside is indicated, expected chromosome numbers are indicated in parentheses. **b)** Alignment of LRRs of domain C to the LRRs of Cf-9 using Clustal Omega. LRRs were predicted using LRRfinder. Numbers indicate total number of LRRs in domain C, putative ligand is indicated for *Hcr9* family members with identified recognition specificities. A, signal peptide, B, N-terminal LRR flanking domain, C, LRR region, D, juxtamembrane region, E, transmembrane domain, F, cytoplasmic tail.

Supplemental data:

Table S1: Rcr3, Pip1, Pip1-like, Rfp1, Pfp1, and Hcr2 homologs identified.

Table S2: *Nicotiana benthamiana* PLCPs.

Table S3: Differences in Peptidase C1a domain of cloned Rcr3 sequences.

Table S4: Previously published PLCP interactions with Avr2.

Table S5: Sequences of synthetic genes used in this study.

Supplemental dataset 1: Complete gels of figures.

Chapter 4

Decoy Engineering: The Next Step in Resistance Breeding

Kourelis, J., van der Hoorn, R.A.L. and Sueldo D. (2016) Decoy engineering: the next step in resistance breeding. *Trends in Plant Science*, **21**, 371-373

Authors

Jiorgos Kourelis[1], Renier A.L. van der Hoorn[1*], and Daniela Sueldo[1,2]

Affiliations

[1]The Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, OX1 3RB Oxford, United Kingdom

[2]Current address: School of Life Sciences, Gibbet Hill Campus, The University of Warwick, CV4 7AL Coventry, United Kingdom

[*]Address correspondence to renier.vanderhoorn@plants.ox.ac.uk.

Author contributions

RALvdH conceived the research. JK developed the storyline with feedback from RALvdH and DS. RALvdH generated the figure with feedback from JK and DS. JK and DS wrote the manuscript with feedback from RALvdH.

Justification

In this commentary we discuss the concept of decoy engineering in order to create novel disease resistance specificities. Conceptually this chapter bridges Chapter 3, where I describe the evolution of Cf-2/Rcr3 and Chapter 5, where I describe the interaction between Cf-2 and Rcr3. Understanding of evolution and molecular interactions is vital for future research aimed at engineering novel disease resistance specificities using these components.

Decoy engineering: the next step in resistance breeding

Jiorgos Kourelis[1], Renier A. L. van der Hoorn[1*], Daniela Sueldo[1]

[1]The Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, Oxford, UK

[*]Correspondence: renier.vanderhoorn@plants.ox.ac.uk (R.A.L. van der Hoorn).

Abstract: The recent finding that decoy engineering can expand the recognition specificity of a plant immune receptor opens a wealth of opportunities for resistance breeding. In this Spotlight we discuss which factors should be considered to successfully translate decoy engineering into crop species.

Keywords: PBS1 // AvrPphB // RPS5 // Disease resistance // Decoy model // Effector // Pseudomonas syringae // Protein engineering

Until recently it was not possible to engineer novel recognition specificities of classical plant immune receptors to completely unrelated effectors. In a recent publication, Kim *et al*. engineered a plant effector target to increase novel recognition specificities by trapping unrelated pathogen-derived proteases in their act (Kim *et al.*, 2016b).

RESISTANCE TO PSEUDOMONAS SYRINGAE 5 (RPS5) is a plant immune receptor of the Nucleotide-binding, leucine-rich repeat (NLR) type which perceives the *Pseudomonas syringae* Type-III effector AvrPphB, a papain-like cysteine protease belonging to the Peptidase C58 family (Shao *et al.*, 2003b). The perception of AvrPphB by RPS5 requires one additional host-derived factor known as AVRPPHB SUSCEPTIBLE 1 (PBS1), which belongs to the Subfamily VII of Receptor-like Cytoplasmic Kinases (RLCK VII). Upon bacterial infection, PBS1, which binds to RPS5 in its pre-activation state, is cleaved by AvrPphB. PBS1 cleavage exposes a five amino acid loop in PBS1 that is believed to activate RPS5, triggering an immune

response characterized by the hypersensitive response (HR), a form of programmed cell death (D. Qi *et al.*, 2014). Interestingly, RPS5-mediated immune signaling requires both PBS1 fragments, and the conformational change induced by cleavage can be mimicked by insertion of five amino acids in the AvrPphB cleavage site (Deyoung *et al.*, 2012). Therefore, perception of AvrPphB follows a mouse-trap mechanism where cleavage of PBS1 (bait) sets off the trap and activates RPS5, triggering immune responses.

RPS5 detects PBS1 cleavage through its concomitant conformational change, irrespective of the presence of AvrPphB. Based on this, Kim *et al*. inferred that an RPS5-mediated immune response against other pathogens could be engineered by exchanging the AvrPphB cleavage sequence with the cleavage sequence of other pathogen-secreted proteases (**Figure 1**). To test this, three PBS1 mutant alleles were generated, containing cleavage sites of three different pathogen-derived proteases: the Type-III effector AvrRpt2 of *P. syringae* ($PBS1^{RCS2}$); the NIa protease of Tobacco Etch Virus (TEV) ($PBS1^{TEV}$); and the NIa protease of Turnip Mosaic Virus (TuMV) ($PBS1^{TuMV}$). As predicted, these PBS1 mutant proteins were specifically cleaved only upon co-expression with the corresponding pathogen-derived protease, triggering RPS5-dependent HR. Importantly, stable expression of $PBS1^{RCS2}$ under its own promotor in Arabidopsis provided resistance against *P. syringae* expressing AvrRpt2 in the absence of RPS2 and RIN4, two proteins required for AvrRpt2-triggered immunity. In addition, these lines recognize AvrPphB, as they still carry the wild-type *PBS1* allele, indicating that expression of *PBS1* mutant alleles does not interfere with the RPS5-dependent perception of AvrPphB by wild-type PBS1. Likewise, Arabidopsis lines stably expressing $PBS1^{TuMV}$ mounted an immune response upon infection with TuMV, which expresses the TuMV-Nla protease. The ease by which novel specificities are introduced by PBS1 modifications makes one wonder if PBS1 homologs might have evolved to confer different resistance specificities.

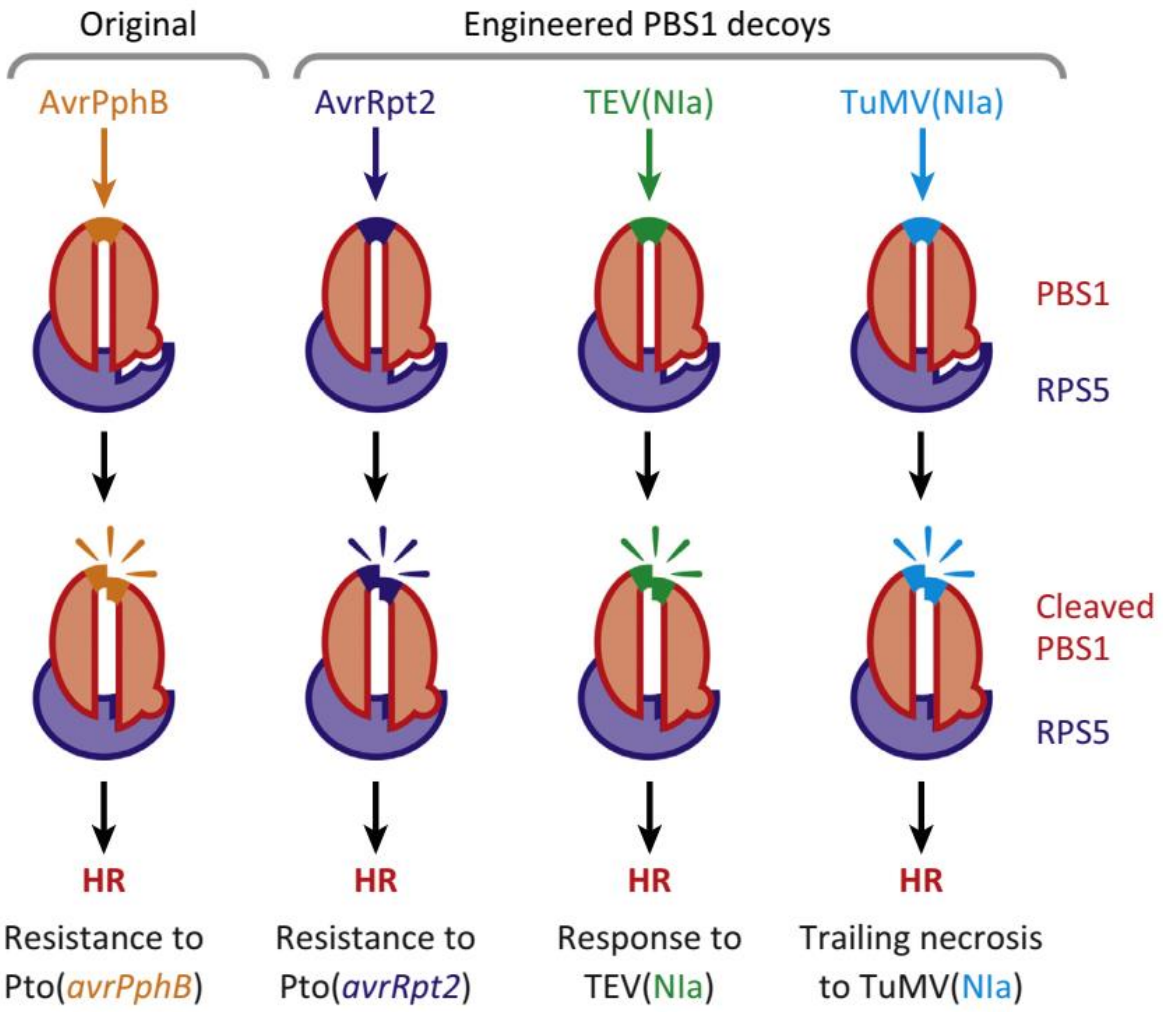


Figure 1. Engineering of the PBS1 decoy by introducing cleavage sites for unrelated proteases results in novel recognition specificities of the RPS5 immune receptor. The native PBS1 (left) was engineered by replacing the AvrPphB cleavage site (orange) by cleavage sites targeted by unrelated pathogen-derived proteases: AvrRpt2 (blue) from *Pseudomonas syringae* (Pto); Nla protease (green) of Tobacco Etch Virus (TEV) and the Nla protease (cyan) of Turnip Mosaic Virus (TuMV).

But there are important considerations when modifying a decoy to engineer novel pathogen recognition specificities. First, for an immune response to be effective and timely, all components involved in perception should localize to the same subcellular compartment. Indeed, whereas the RPS5/PBS1 complex localizes to the plasma membrane (PM), the TuMV-Nla protease mainly resides in the nucleus (Qi *et al.*, 2012). Hence, though cleavage of $PBS1^{TuMV}$ by TuMV-NIa activates RPS5, the induction of the immune response is delayed to a late stage of infection, when enough TuMV-Nla protein has accumulated in the cytoplasm

and the virus has spread. As a result of this, Arabidopsis lines overexpressing PBS1^{TuMV} suffered from trailing necrosis upon infection with TuMV, a phenomenon wherein HR trails the spread of the virus. Kim *et al*. suggest targeting RPS5/PBS1 to the same subcellular location as TuMV-NIa protease to circumvent this issue. However, several studies have shown that subcellular localization is important for the proper activation of multiple NLR proteins, including RPS5 (Qi *et al.*, 2012; Gao *et al.*, 2011). Targeting RPS5 to a different subcellular compartment could impair its downstream signaling. Second, there must be constraints as to how much PBS1 can be altered without triggering auto-immune responses. For example, engineering PBS1 to recognize a pathogen-derived protease with a similar substrate specificity as a host protease would inevitably trigger auto-immunity. In addition, a slight alteration of the PBS1 sequence may be enough to trigger HR. For instance, inserting five amino acids in the AvrPphB cleavage site triggers autoimmune responses (Deyoung *et al.*, 2012). Finally, in case an effector target has an important physiological function, modifications of such an effector target should not impair its original function or interfere with the wild-type effector target when both alleles are present.

PBS1-based protease traps may be easily extended to perceive proteases from other pathogens. Pathogen-derived proteases are important effectors, and are produced by many unrelated pathogens. Given the fact that these effector proteases alter the host cell, they must have a specificity or spatial/temporal regulation that makes their activity distinct from endogenous host proteases. This specificity can be exploited while building other PBS1-like protease traps. Interestingly, plants have evolved additional mechanisms to recognize pathogen-derived proteases through their activity. RIN4, for instance, is an effector target whose cleavage by AvrRpt2 triggers immunity mediated by the NLR RPS2 (Gao *et al.*, 2011). It would be interesting to explore whether replacing the AvrRpt2 cleavage site in RIN4 for that of AvrPphB activates RPS2-mediated HR. However, we predict that engineering RIN4 is more challenging

than PBS1, as the AvrRpt2 cleavage site overlaps with a region that is crucial for other immune-related functions (Li *et al.*, 2014). Several pathogen-derived apoplastic proteases, such as *Xanthomonas campestris* argC and *Clavibacter michiganensis* Chp-7 and ChpG are perceived in an activity-dependent manner in *Brassicaceae* and *Nicotiana* subspecies, respectively (Cheng *et al.*, 2015; Lu *et al.*, 2015). Identification of the molecular components underlying perception of these extracellular proteases could facilitate engineering the perception of apoplastic pathogen-derived proteases, mediating recognition in subcellular compartments other than the cytoplasm.

In contrast to effectors with other enzymatic activities, the identification of preferential cleavage motifs of effector proteases can be relatively straightforward, for example by using by simple proteomic techniques such as PICS (Proteomic Identification of protease Cleavage Sites) (Schilling and Overall, 2008). Furthermore, only a few amino acids in the host target need to be substituted, as protease specificity is usually determined by short peptide sequences, thereby reducing the chance of unintended side effects. Engineered decoy proteins can be quickly screened by transient expression for specific cleavage by effector proteases and triggering HR to facilitate pre-selection before generating stable lines.

Also other decoys can be engineered to change their specificity. Apart from effector target cleavage, inhibition, phosphorylation and other post-translational modifications of guarded effector targets have been reported. For example, tomato Cf-2 is a transmembrane immune receptor that perceives the interaction of the apoplastic cysteine protease Rcr3 with Avr2, a protease inhibitor produced by the fungus *Cladosporium fulvum* (Dixon *et al.*, 2000). It would be interesting to investigate if the concept of decoy engineering can be extended to this perception mechanism. Likewise, the promotors of the so-called executor genes have been successfully engineered to trap unrelated TAL (Transcription-activator like) effectors produced by *Xanthomonas spp* (Zeng *et al.*, 2015).

In conclusion, the work of Kim *et al.* provides a proof-of-principle for new crop protection strategies, by showing that pathogen-derived proteases can be tricked into cleaving PBS1, thus turning this host kinase into a versatile protease trap.

Acknowledgements. The authors are financially supported by ECR Consolidator grant 616449 'GreenProteases', the University of Oxford and The Clarendon Fund.

Chapter 5

Induction of the hypersensitive response by Rcr3 resides in few Rcr3-specific amino acid residues

Kourelis, J., Mattinson, O., Krauter, S., Kahlon, P.S., O'Grady, R., Royston, S., Demir, F., Huesgen., P., van der Hoorn, R.A.L. (2018) Induction of the hypersensitive response by Rcr3 resides in few Rcr3-specific amino acid residues. *In preparation*

Authors

Jiorgos Kourelis, Oliver Mattinson, Sonja Krauter, Parvinderdeep S. Kahlon, Ruby O'Grady, Samantha Royston, Fatih Demir, Pitter Huesgen, Renier A. L. van der Hoorn

Author contributions

Conceptualization: JK, RvdH; Formal analysis: JK; Investigation: JK, and OM, SK, ROG, SR and PSK under supervision of JK; PICS analysis: FD, PH; Funding acquisition: RvdH; Writing: JK, RvdH.

Justification

In this chapter I describe the molecular mechanism of the Cf-2/Rcr3 perception mechanism. I show that it is possible to engineer Pip1 to take over the role of Rcr3

Induction of the hypersensitive response by Rcr3 resides in few Rcr3-specific amino acid residues

Jiorgos Kourelis[1], Oliver Mattinson[1], Sonja Krauter[1], Parvinderdeep S. Kahlon[1,2], Ruby O'Grady[1,3], Samantha Royston[1], Fatih Demir[4], Pitter Huesgen[4], Renier A. L. van der Hoorn[1*]

[1]Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, South Parks Road, OX1 3RB Oxford, UK

[2]Current address: Lehrstuhl für Phytopathologie, Technische Universität München, Arcisstraße 21, 80333 München, Germany

[3]Current address: The Sainsbury Laboratory, Norwich Research Park, Norwich, UK

[4]Central Institute for Engineering, Electronics and Analytics (ZEA), Analytics (ZEA-3), Research Centre Jülich, Wilhelm-Johnen-Str., 52428 Jülich, Germany

[*]Correspondence: renier.vanderhoorn@plants.ox.ac.uk (R.A.L. van der Hoorn).

Abstract

The tomato resistance gene *Cf-2* provides immunity to races of the fungal pathogen *Cladosporium fulvum* producing the secreted small cysteine-rich effector Avr2. *Cf-2*-dependent recognition of Avr2 triggers a hypersensitive response (HR) which also requires tomato *Rcr3*, which encodes for a papain-like cysteine protease. We previously showed that the genetic interaction of *Cf-2* with Rcr3, and the interaction of Rcr3 with Avr2, can be genetically separated. To identify residues required for *Cf-2*-dependent Avr2 recognition, we first engineered the *Nicotiana benthamiana* and eggplant (*Solanum melongena*) Rcr3 homologs to trigger Cf-2/Avr2-dependent HR. Using a fragment-swapping approach we next engineered the Rcr3-like protease Pip1 to trigger Cf-2/Avr2-

dependent HR. Further mutagenesis studies revealed that approximately 12 Rcr3-specific residues are required for Cf-2/Avr2-dependent HR. Rcr3 and Pip1 have been separated by over 60 million years of evolution and have distinct substrate specificities. The identified residues in Rcr3 important for HR induction reside around the substrate-binding groove, but are distinct from those previously identified to be important for inhibition by Avr2. These data support a model in which Cf-2 and Rcr3 interact directly, and binding of Avr2 triggers a conformational rearrangement leading to the downstream immune signalling.

Keywords: Solanaceae // Plant Innate Immunity // R-gene transfer // Rcr3 // Pip1 // Avr2 // Cladosporium fulvum // Nicotiana benthamiana // Tomato

Introduction

Plants rely on resistance (*R*) genes for innate immunity to pathogens. Although the molecular identity of many *R* genes has been uncovered, the molecular mechanisms by which they function are only understood for a few (Kourelis and van der Hoorn, 2018). Understanding these molecular mechanisms is crucial for the deployment and rational engineering of *R* genes.

Several *R* genes have been cloned which confer resistance to the leaf mould causing fungal pathogen *Cladosporium fulvum* (syn. *Passalora fulva*) in tomato (*Solanum lycopersicum* L.). All these genes encode for receptor-like proteins (RLPs) with extracellular leucine-rich repeat (LRR) domains.

One such *R* gene is *Cf-2*, which was introgressed from the currant tomato (*Solanum pimpinellifolium*) (Dixon *et al.*, 1996). *Cf-2* confers recognition of the secreted small cysteine-rich effector Avr2 (Luderer *et al.*, 2002). This recognition uniquely requires tomato *Rcr3*, which encodes for a papain-like cysteine protease (Dixon *et al.*, 2000).

Avr2 acts as a protease inhibitor (Rooney *et al.*, 2005). *Cf-2* tomato lines which lack *Rcr3* do not display any obvious phenotypes (Dixon *et al.*, 2000). This indicates that it is not the inhibition of Rcr3 activity which is recognized by *Cf-2*, but rather the physical interaction of Rcr3 with Avr2. Further studies using mutant Avr2 protein showed that the inhibition of Rcr3 by Avr2 positively correlates with the induction of *Cf-2*-dependent HR (van't Klooster *et al.*, 2011). Furthermore, Rcr3 homologs with a N194D substitution cannot be inhibited by Avr2, and do not trigger *Cf-2*-dependent recognition of Avr2 (Shabab *et al.*, 2008; Hörger *et al.*, 2012). However, inhibition of Rcr3 by Avr2 is not sufficient to trigger *Cf-2*-dependent recognition because some Rcr3 homologs which can be inhibited by Avr2 do not trigger a *Cf-2*-dependent HR upon this interaction (chapter 3, this thesis). This indicates that the residues required for the interaction with Avr2, and those required for *Cf-2*-dependent Avr2 recognition are distinct. Identifying the residues required for *Cf*-2-dependent Avr2 recognition could lead to a more detailed understanding of the molecular mechanism by which *Cf-2* recognizes the interaction of Avr2 with Rcr3.

Here we investigated which residues in Rcr3 are required to trigger a *Cf-2*-dependent HR upon Avr2 inhibition. We also engineered the Rcr3-like protease Pip1 to trigger *Cf-2*-dependent Avr2 recognition. This work provides exciting new insights into the molecular mechanism of a guarded decoy and its receptor. Additionally, this guardee-swapping could provide a novel strategy for the future engineering of this molecular mechanism for an altered recognition spectrum.

Results

Distant Rcr3 homologs can be engineered for Cf-2-dependent Avr2 recognition

We previously showed that while the resurrected *Nicotiana benthamiana* Rcr3 ($NbRcr3_{res}$) with a G194N substitution can be inhibited by Avr2, it cannot trigger HR upon co-expression with Cf-2 and Avr2. We realized that all *Nicotiana* Rcr3 homologs were missing three residues

(D328/P329/S330) present in *Capsicum* and *Solanum* Rcr3 homologs. We reasoned that these residues may be required for Cf-2/Avr2-dependent HR.

Activity-based protein profiling with the MV201 probe, a fluorescent derivative of the general PLCP inhibitor E-64 (Richau *et al.*, 2012), was used to test the activity of $NbRcr3_{res}$ mutants in which we added the residues D328/P329/S330 ($NbRcr3_{res}$(DPS)). Both the $NbRcr3_{res}$(DPS) and $NbRcr3_{res}$(G194N/DPS) mutants are active proteases (**Figure 1a**), but only $NbRcr3_{res}$(G194N/DPS) is inhibited by Avr2 (**Figure 1a**). Only the co-expression of the $NbRcr3_{res}$(G194N/DPS) mutant with Cf-2 and Avr2 triggers HR in *N. benthamiana*, and not $NbRcr3_{res}$(DPS) or $NbRcr3_{res}$(G194N) (**Figure 1b**). This data demonstrate that D328/P329/S330 are required for HR induction but not for Avr2 inhibition. Similarly, the addition of residues D328/P329/S330 in the *N. sylvestris* and *N. tabacum* Rcr3 homologs results in a HR upon co-expression with Cf-2 and Avr2 in *N. benthamiana*, dependent on an additional K194N mutation which allows for Avr2 inhibition (**Figure S1**).

We next took a closer look at the eggplant (*Solanum melongena* L.) Rcr3 (SmeRcr3). SmeRcr3 can be inhibited by Avr2 but triggers a very weak HR upon co-expression with Avr2 and Cf-2 (chapter 3, this thesis). SmeRcr3 contains N328/P329/G330, as present in potato Rcr3 homologues which can trigger a HR upon co-expression with Cf-2 and Avr2 in *N. benthamiana*. This indicates that there is an independent variation which results in an inability to interact with *Cf-2*. We used an untargeted approach in which we replaced large fragments of SmeRcr3 with the corresponding $Rcr3^{pim}$ sequence (**Figure S2a**). Of the tested hybrid proteases only SmeRcr3 containing fragment 4 of $Rcr3^{pim}$ was able to trigger HR upon co-expression with Cf-2 and Avr2 (**Figure S2b**). This focussed our attention to an aspartic acid residue in SmeRcr3 (D244). All Rcr3 homologs that can trigger HR upon co-expression with Cf-2 and Avr2 contain a proline at this position. The SmeRcr3(D244P) mutant is not affected in Avr2 inhibition (**Figure 1c**). However, unlike the wild-type SmeRcr3, the SmeRcr3(D244P) mutant

can trigger an efficient HR upon co-expression with Cf-2 and Avr2 (**Figure 1d**), demonstrating that P244 like D328/P329/S330 is required for HR induction but not for Avr2 inhibition. The residues P244, and D328, P329, and S330 are in close proximity to each other residing near the S2 substitute site on the right-hand lobe of Rcr3 (**Figure 1e**).

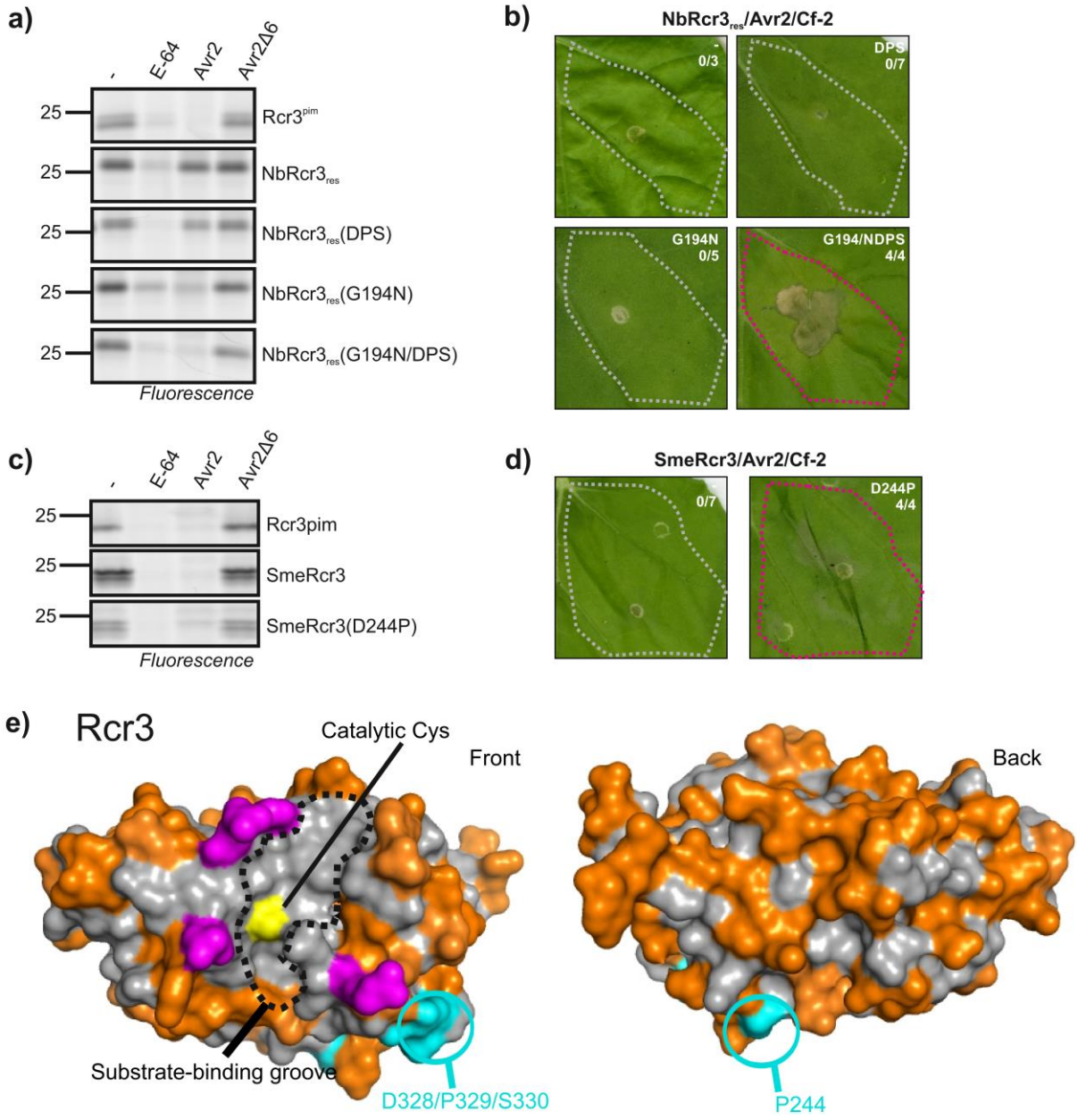


Figure 1: *N. benthamiana* and eggplant Rcr3 can be engineered to trigger a Cf-2/Avr2-dependent HR. **a)** Activity-based protein-profiling of Papain-like cysteine proteases. $Rcr3^{pim}$ and NbRcr3 mutants were co-expressed with silencing inhibitor P19 by agroinfiltration of *N. benthamiana* leaves. Apoplastic fluids were isolated at 2 days-post-infiltration (dpi), pre-incubated for 45 min with and without 100 µM E-64, 1 µM Avr2 or Avr2Δ6, and labelled for 3 hours with 0.2 µM MV201. Samples were separated on SDS-PAGE gels and scanned for

Figure 1. (Continued). fluorescence. **b)** Co-infiltration of *A. tumefaciens* containing binary vectors for the expression of Cf-2, Avr2, and either $NbRcr3_{res}$, $NbRcr3_{res}$(DPS), $NbRcr3_{res}$(G194N), or $NbRcr3_{res}$(G194N/DPS). Only $NbRcr3_{res}$(G194N/DPS) triggers a HR in *N. benthamiana*. *A. tumefaciens* containing different binary vectors were combined at a final OD_{600} = 0.25 each in a 1:1:1 ratio. The silencing inhibitor P19 was present in the backbone of all binary vectors. Pictures were taken at 5 dpi. Numbers indicate the number of leaves showing HR. **c)** Activity-based protein-profiling of Papain-like cysteine proteases. $Rcr3^{pim}$, SmeRcr3, and SmeRcr3(D244P) were produced and labelled as in (**a). d)** Transient expression of the SmeRcr3(D244P) mutant with Cf-2 and Avr2 triggers a HR in *N. benthamiana*. Co-infiltration of *A. tumefaciens* containing different binary vectors was done as in (**b**). **e)** Homology model of $Rcr3^{lyc}$ based on 1s4v (Ilyas *et al.*, 2015). Dotted line marks the substrate-binding groove, the catalytic cysteine is indicated in yellow, Magenta indicates the H148, R151, N194, and Q284 residues previously identified to be important for the interaction with Avr2. Orange are polymorphic residues in Rcr3 homologs which do not affect *Cf-2*-dependent Avr2 recognition. Cyan indicates residue P244, and D328, P329, S330 required to confer *Cf-2*-dependent Avr2 recognition to eggplant Rcr3 and *N. benthamiana* Rcr3, respectively.

A screen of Rcr3/Pip1 hybrids reveals three distinct regions of Rcr3 required for Cf-2-dependent Avr2 recognition

To further identify Rcr3-specific residues required for Cf-2/Avr2-dependent HR we decided to investigate functional differences with the Rcr3-like protease Pip1. *Rcr3* and *Pip1* are conserved in Solanaceae and have been separated by at least 60 million years of evolution (chapter 3, this thesis). Both Rcr3 and Pip1 can be efficiently inhibited by Avr2 (Shabab *et al.*, 2008), facilitating the identification of residues required for Cf-2/Avr2-dependent HR.

Rcr3 and Pip1 are produced as preproproteins, containing a signal peptide targeting them to the apoplast, an inhibitor I29-containing pro-domain, and a peptidase C1 domain. Rcr3 detected in the apoplast is maturated and lacks the pro-domain (Rooney *et al.*, 2005). Therefore, residues which could confer Cf-2/Avr2-dependent HR in Rcr3 are likely in the peptidase C1 domain. The peptidase C1 domain of $Rcr3^{pim}$ and Pip1 are 54.8% identical (**Table S1**). Using the same fragment-swapping approach as used for SmeRcr3, in which we divided the peptidase C1 domain of Rcr3 in six fragments, we tested 57 Rcr3/Pip1 hybrid proteases containing the Pip1 pro-domain (**Figure S3**, **Table S1**). These hybrid proteases were expressed in *N. benthamiana* by agroinfiltration. Of the 57 tested hybrid proteases, 33 were active proteases as determined

by activity-based protein profiling of apoplastic fluids isolated from agroinfiltrated *N. benthamiana* with the MV201 probe (**Figure S4a**). It is likely that the other 24 Rcr3/Pip1 hybrids are unstable due to a disruption of the hydrophobic core of the protease. In addition, cysteines may be in slightly different positions in the Rcr3/Pip1 hybrids preventing them from forming disulphide-bridges. To establish whether these 33 hybrid proteases could trigger Cf-2/Avr2-dependent HR, we isolated apoplastic fluids from *N. benthamiana* transiently expressing these hybrid proteases and injected this together with 1 µM recombinant Avr2 produced in *E. coli* in the tomato MM-*Cf-2 rcr3-3* line which does not produce a functional Rcr3. Out of the 33 active proteases, eight can trigger *Cf-2*-dependent HR upon co-infiltration with 1µM Avr2 in tomato MM-*Cf-2 rcr3-3* lines (**Figure S4b**).

All eight Rcr3/Pip1 hybrids which can trigger Cf-2/Avr2-dependent HR can also be inhibited by Avr2 (**Figure 2a**). 22 the remaining 25 active Rcr3/Pip1 hybrids can be inhibited by Avr2 (**Figure S2c**). The three hybrids which cannot be inhibited by Avr2 contain fragment 3 of Pip1 containing E194, and fragment 5 of Rcr3 containing Q284. We previously showed that the combination of E194 and Q284 prevents Avr2 inhibition (chapter 3, this thesis). Pip1 containing fragments 3, 5, and 6 from Rcr3 can trigger *Cf-2*-dependent Avr2 recognition (**Figure 2b**; construct 12). The peptidase C1 domain of this Rcr3/Pip1 hybrid is 75.9% identical to Pip1. The variant residues in these fragments mostly surround the substrate-binding groove on the right-hand lobe of Rcr3 (**Figure 2c**), where also the residues required to confer Cf-2/Avr2-dependent HR in *N. benthamiana* and eggplant Rcr3 localize.

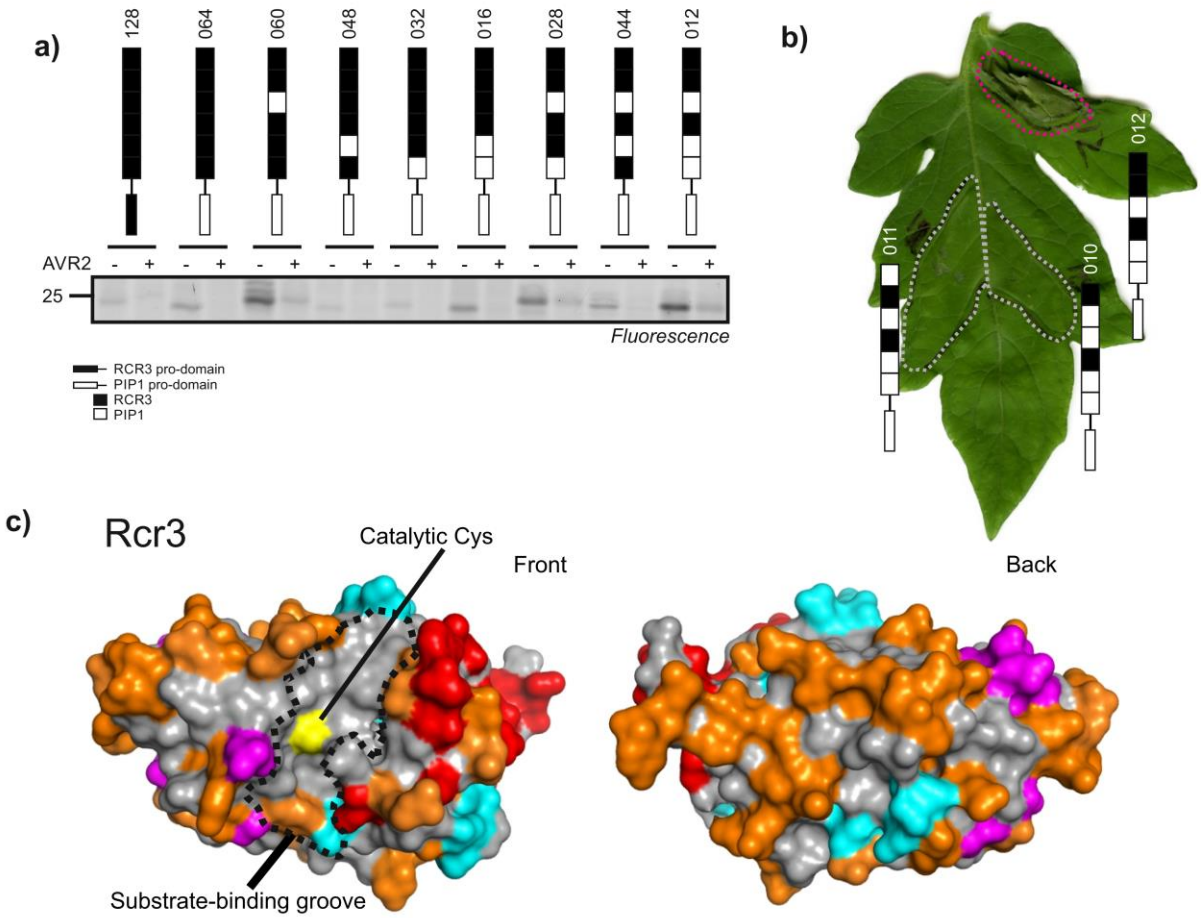


Figure 2: A series of Rcr3/Pip1 hybrid proteins reveal three distinct regions in Rcr3 required for Cf-2/Avr2-dependent HR. **a)** Activity-based protein-profiling of Papain-like cysteine proteases. Rcr3/Pip1 fragment-swapped hybrids were co-expressed with silencing inhibitor P19 by agroinfiltration of *N. benthamiana* leaves. Apoplastic fluids were isolated at 5 days-post-infiltration (dpi), and incubated for 45 min with and without 500 nM Avr2 and labelled for 5 hours with 2 µM MV201. Samples were separated on SDS-PAGE gels and scanned for fluorescence. **b)** Pip1 containing fragment 3, 5, and 6 of Rcr3 can trigger Cf-2/Avr2-dependent HR. Apoplastic fluids containing Rcr3/Pip1 hybrids produced as above were co-injected with 1 µM *E. coli* produced recombinant Avr2 in tomato MM-*Cf-2 rcr3-3* plants. Pictures taken at 5 dpi. **c)** Homology model of $Rcr3^{lyc}$ based on 1s4v (Ilyas *et al.*, 2015). Dotted line marks the substrate-binding groove, the catalytic cysteine is indicated in yellow. Grey indicates identical residues between Rcr3 and Pip1, whereas orange indicates divergent residues not important for the Cf-2/Avr2-dependent HR. Variant residues in fragments 3, 5, and 6 potentially involved in Cf-2/Avr2-dependent HR are indicated in magenta, red, and cyan, respectively.

Mutagenesis in Rcr3/Pip1 hybrid proteases pinpoints 12 residues required for Cf-2/Avr2-dependent HR

To further identify critical residues in Rcr3 required to trigger Cf-2/Avr2-dependent HR we took the Pip1 hybrid protease containing fragments 3, 5, and 6 from Rcr3, and in each fragment made a series of mutants.

In fragment 3 there are 12 variant residues between Pip1 and Rcr3 which may be important to confer Cf-2/Avr2-dependent HR (**Figure 3a**). This fragment contains residue N194, which was previously shown to be important for Avr2 inhibition (Shabab *et al.*, 2008). We tested nine mutants in which we had mutated various combinations of these variant residues. Surprisingly, all mutant proteases were able to trigger Cf-2/Avr2-dependent HR (**Table S2**). Fragment 3 is important for Avr2 inhibition, as a hybrid Rcr3/Pip1 protease containing only fragment 5 and 6 of Rcr3 is not inhibited by Avr2 (**Figure S4c**), but this hybrid also accumulates to much lower level than those containing fragment 3 (**Figure S4a**). This indicates that fragment 3 is required for stability and Avr2 inhibition. The different sensitivity to Avr2 can be explained by E194N. We expect that some of the mutants which could trigger Cf-2/Avr2-dependent HR, but lacked E194N, contain the mutations which allow the protease to accumulate to higher levels. Looking at the residues (NNG208N) common between the mutants which lack E194N, this might be due the deletion of two additional residues not present in Rcr3. Additionally, this deletion could alter the conformation of distant residues. We therefore expect that E194N, and the deletion of these two residues to be sufficient to ensure Avr2 inhibition and accumulation.

Fragment 5 contains 14 variant residues between Pip1 and Rcr3 which may be important to confer Cf-2/Avr2-dependent HR (**Figure 3b**). Additionally, this fragment contains Q284, which we previously showed was important for Avr2 inhibition (chapter 3, this thesis), and A282 which is likely required for SmeRcr3 to trigger HR upon co-expression with Cf-2 and Avr2. We tested ten mutant proteases in which we had mutated various combinations of

putative surface-localized variant residues (**Table S3**). Four residues were common to all mutants which could trigger HR upon co-expression with Cf-2 and Avr2 (**Figure 3b**, **Table S3**; N266S/D267Q/H270Q/M271F). Surprisingly this required either of two other regions: one containing A282 (S282N/N284A/S285D) or one containing a deletion of one residue (SE297D/D299E/T301Q). These resides all surround the substrate-binding groove on the right-hand lobe of Rcr3.

Finally, fragment 6 contains 15 variant residues which may be important to confer Cf-2/Avr2-dependent HR (**Figure 3c**). Additionally, this fragment contains D328, P329, and S330 which are required for NbRcr3 to trigger HR upon co-expression with Cf-2 and Avr2. To further pinpoint residues in this fragment we first divided the variant residues into two halves. The seven variant residues between Rcr3 and Pip1 in the C-terminal half of this fragment are sufficient to trigger Cf-2/Avr2-dependent HR (**Figure 3c**, **Table S4**). Another mutant protease containing only five variant residues can trigger Cf-2/Avr2-dependent HR (**Figure 3c**, **Table S4**; D329P/G334D/V338M/A339S/F341Y).

The residues required to trigger Cf-2/Avr2-dependent HR mostly surround the substrate-binding groove on the right-hand lobe of the protease (**Figure 3d**; red residues). Currently we are further investigating these residues and assembling them in a single mutant Pip1 protease which can trigger Cf-2/Avr2-dependent HR.

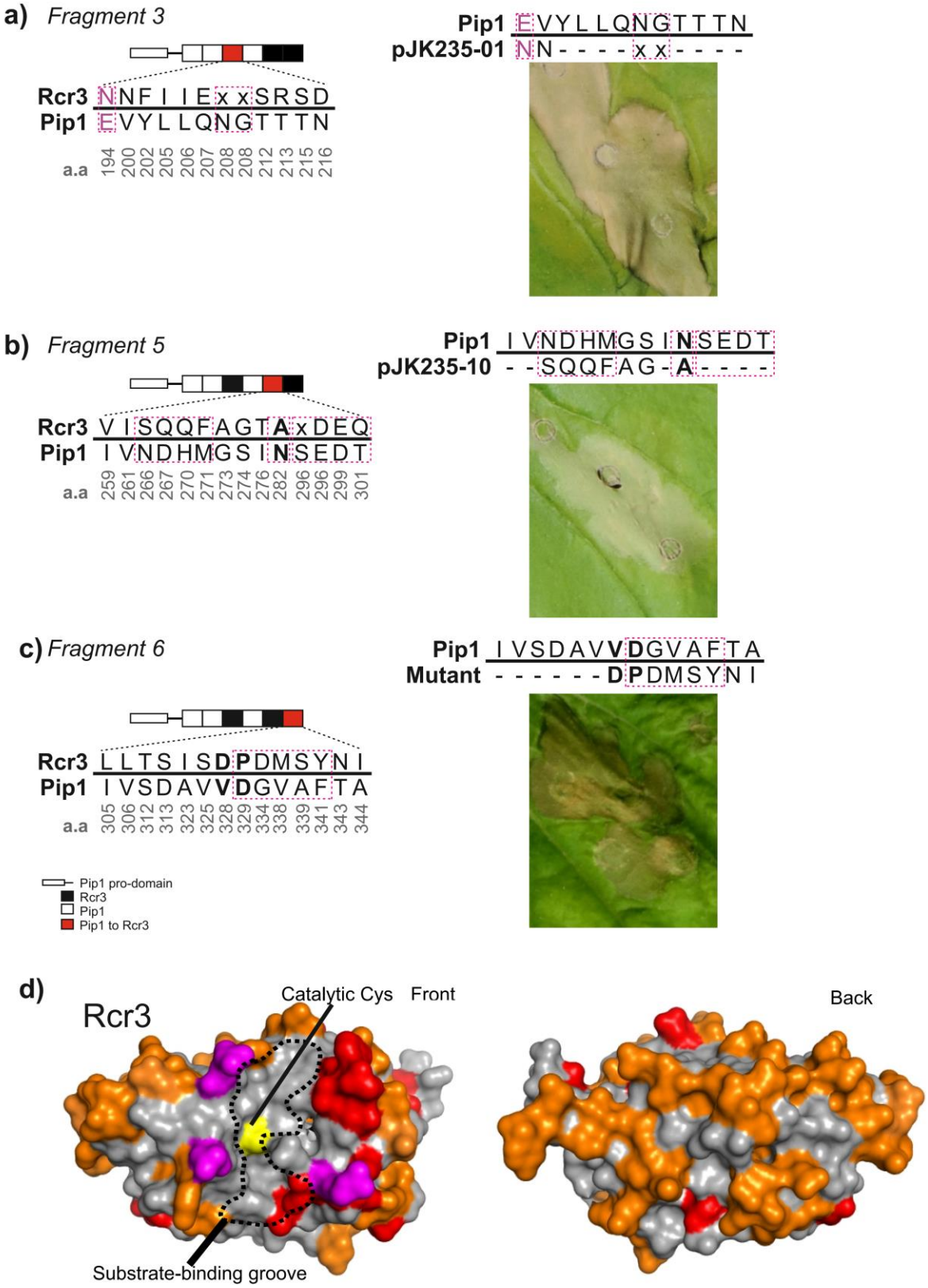


Figure 3: Further mutagenesis of Rcr3/Pip1 hybrid proteases. **a)** Variant residues between Rcr3 and Pip1 in fragment 3 required to trigger Cf-2/Avr2-dependent HR. An example is shown. **b)** Variant residues between Rcr3 and Pip1 in fragment 5 required to trigger Cf-2/Avr2-dependent HR. An example is shown. **c)** Variant residues between Rcr3 and Pip1 in fragment 6 required to trigger Cf-2/Avr2-dependent HR. An example is shown. **a-c)** Residues

Figure 3. (Continued). important for triggering Cf-2/Avr2-dependent HR in Rcr3 homologs are indicated in bold, residues important for Avr2 inhibition are indicated in magenta. Residues in boxes are residues common between mutant hybrids able to trigger Cf-2/Avr2-dependent HR. *A. tumefaciens* containing binary vectors for the expression of Cf-2, Avr2, and Rcr3/Pip1 mutant hybrids were combined at a final OD_{600} = 0.25 each in a 1:1:1 ratio and co-infiltrated in *N. benthamiana*. The silencing inhibitor P19 was present in the backbone of all binary vectors. Pictures were taken at 5 dpi. **d)** Homology model of $Rcr3^{lyc}$ based on 1s4v (Ilyas *et al.*, 2015). Dotted line marks the substrate-binding groove, the catalytic cysteine is indicated in yellow. Grey indicates identical residues between Rcr3 and Pip1, whereas orange indicates divergent residues not important for the Cf-2/Avr2-dependent HR. Variant residues in fragments 3, 5, and 6 that might be required for Cf-2/Avr2-dependent HR are indicated in red. Residues important for Avr2 interaction are indicated in magenta.

Rcr3 and Pip1 have distinct substrate specificity

Finally, we investigated whether Rcr3 and Pip1 have distinct substrate specificities. To do this we used a modified version of proteomic identification of protease cleavage sites (PICS) to identify substrate specificity (Biniossek *et al.*, 2016) (**Figure 4a**). Briefly, we generated an *E. coli* Trypsin/Lys-C peptide library which we digested with recombinant C-terminal 6xHis-tagged $Rcr3^{pim}$ and Pip1 purified from *Pichia pastoris*. Papain-like cysteine proteases display selectivity in the S2 substrate pocket, which binds the P2 residue in the substrate. Rcr3 has preference for P2 proline (P), methionine (M), valine (V) or leucine (L) (**Figure 4b**). By contrast, Pip1 prefers P2 leucine (L) or phenylalanine (F) (**Figure 4c**). This indicates that these proteases have distinct substrate specificities, and therefore likely have distinct substrates. Whether substrate specificity is required for Cf-2/Avr2-dependent HR is something we are currently investigating.

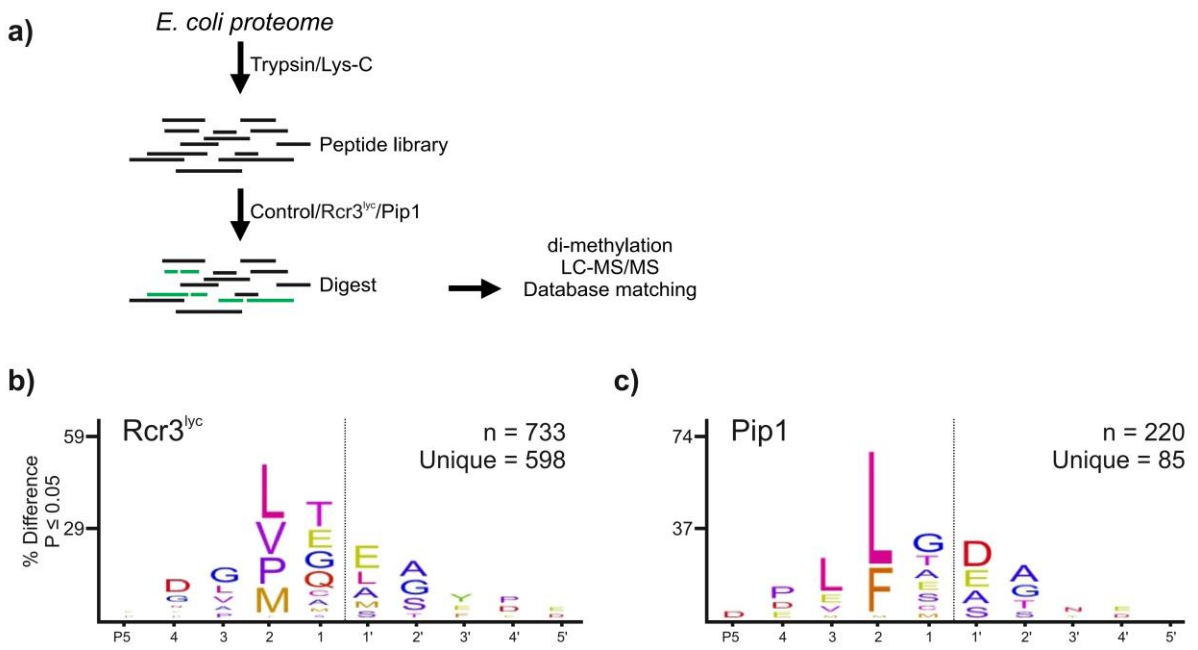


Figure 4: Rcr3 and Pip1 have distinct substrate specificity. **a)** Workflow of the modified version of proteomic identification of cleavage sites (PICS). A peptide library made using Trypsin/Lys-C from an *E. coli* proteome was digested with recombinant C-terminal 6xHis-tagged Rcr3^{pim} and Pip1 purified from *Pichia pastoris*. This digested peptide library was di-methylated, mixed in equal amounts and sent for LC-MS/MS. A custom PERL script was used to filter semi-specific peptides and through database-lookup determine the corresponding prime and nonprime sequence. **b)** and **c)** show sequence logos showing residues preferred by Rcr3^{lyc} (**b**) and Pip1 (**c**) upstream (P1 to P5) and downstream (P1′ to P5′) of the cleavage site, as analysed by PICS. 733 cleaved peptides were identified for Rcr3^{lyc} of which 598 were unique to Rcr3^{lyc}, and 220 cleaved peptides were identified for Pip1 of which 85 were unique to Pip1. 135 cleaved peptides were identified in both Rcr3^{lyc} and Pip1-treated peptide library.

Discussion

Here we identified Rcr3-specific residues required for Cf-2/Avr2-dependent HR. The addition of three residues (D328/P329/S330) to the resurrected *N. benthamiana* Rcr3, the *N. sylvestris* Rcr3, or the *N. tabacum* Rcr3, together with the G194N or K194N substitution reconstitutes Cf-2/Avr2-dependent HR. While the G194N and K194N substitutions are required for Avr2 inhibition, D328/P329/S330 are only required for recognition and are missing in all *Nicotiana* Rcr3 homologs. Likewise, the eggplant Rcr3 can also be engineered to trigger Cf-2/Avr2-dependent HR, by mutating D244P. This residue is does not affect the interaction with Avr2. A series of hybrids between Rcr3 and Pip1 indicated three regions in Rcr3 important for Cf-

2/Avr2-dependent HR. Further fine-mapping of residues important for this HR indicated that they mostly reside on the right-hand lobe of Rcr3 surrounding the substrate-binding groove.

These data lead us to propose a current working model in which Cf-2 and Rcr3 interact directly (**Figure 5**). Binding of strong inhibitors of Rcr3, such as Avr2, triggers a conformational rearrangement with leads to downstream signalling. We have thus far been unable to establish a co-immunoprecipitation of Cf-2, Rcr3, and Avr2. This does not exclude their interaction under natural conditions, and may reflect experimental difficulties of working at apoplastic conditions highly enriched for hydrolases (Kourelis *et al.*, 2018) (Chapter 6, this thesis).

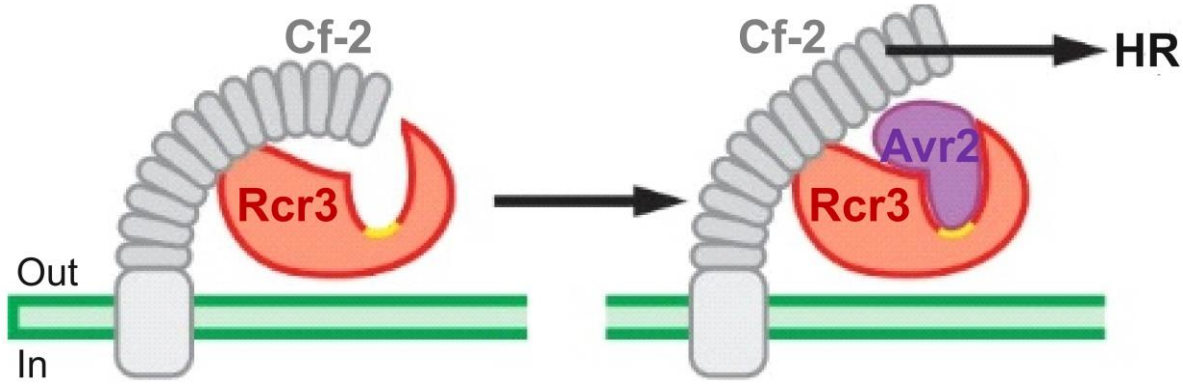


Figure 5: Model of the molecular mechanism of Cf-2/Rcr3-dependent Avr2 recognition. Based on our data we propose that Rcr3 interacts directly with the extracellular leucine-rich repeats of Cf-2, and that binding of strong inhibitors such as Avr2 triggers a conformational rearrangement ultimately resulting in a HR.

Previously it was shown that the RPS5/PBS1 recognition spectrum can be engineered by subtle changes in PBS1 (Kim *et al.*, 2016a). It is possible that guardee-swapping could provide an alternative route to an altered recognition spectrum. Indeed, in nature some *R* genes expand their recognition spectra by interacting with different decoys. It seems that *Cf-2* recognizes the strong interaction of inhibitors with Rcr3. For example, tomato lines containing *Cf-2* and the cultivated tomato homolog of *Rcr3*, *Rcr3*^{lyc}, display an autonecrotic response (Krüger *et al.*, 2002). This autonecrosis is suppressed upon co-introduction of the currant tomato *Rcr3* homolog, *Rcr3*^{pim} (Krüger *et al.*, 2002). It is likely that Rcr3^{lyc}, but not Rcr3^{pim}, interacts with

an additional unknown tomato-derived factor, thereby triggering autonecrosis. Furthermore, the interaction of Rcr3 with the potato cyst nematode (*Globodera rostochiensis*) effector GrVAP1 was reported to trigger a *Cf-2*-dependent HR (Lozano-Torres *et al.*, 2012). However, not all inhibitors of Rcr3 trigger *Cf-2*-dependent recognition. For example, the *Pseudomonas syringae* pv. *tomato* DC3000 chagasin-like effector Cip1, as well as the *Phytophthora infestans* cystatin-like effectors EPIC1 and EPIC2B can inhibit Rcr3, but they do not trigger *Cf-2*-dependent recognition (Shindo *et al.*, 2016; Song *et al.*, 2009). This is probably because these are weak inhibitors of Rcr3 (Kaschani *et al.*, 2010; Shindo *et al.*, 2016). None of the Pip1/Rcr3 hybrids which could trigger *Cf-2*-dependent Avr2 recognition were able to trigger *Cf-2*-dependent recognition of either Cip1, EPIC1, or EPIC2B (**Figure S4b**). However, it is possible that these hybrids could recognize other pathogen components.

In conclusion, the on-going further fine-mapping of the residues within Rcr3 required for *Cf-2*-dependent Avr2 recognition is expected to result in a better understanding of the molecular mechanism of this recognition. Additionally, we propose that guardee-swapping may provide an alternative route for the generation of synthetic *R* genes with altered recognition spectra.

Material & Methods

Fragment-swapped hybrid cloning

For fragment-swapping the nucleotide region encoding for the Rcr3^{pim} and Pip1 peptidase C1 domain were divided into six arbitrary fragments and ordered through gene-synthesis (Biomatik; pJK103-pJK107, pJK111-pJK115, pJK185, pJK186; **Table S5**). These fragments, together with a fragment containing the nucleotides encoding for the Pip1 pro-domain but lacking the signal peptide (pJK110; **Table S5**) were combined in all 64 possible combinations with pICH41264 (Addgene #47993 (Weber *et al.*, 2011)) in a BpiI Golden Gate reaction (**Figure S3**, **Table S5**). These fragment-swapped level 0 modules were subsequently combined

with pJK001 and pJK002 (Grosse-Holz *et al.*, 2018), and pICH51288 and pICH41414 (2x35S and 35S terminator; Addgene #50269 and #50337 (Engler *et al.*, 2014)) in a BsaI Golden Gate reaction to generate binary vectors containing the fragment-swapped Rcr3/Pip1 hybrids driven by the double 35S CaMV promoter and targeted to the apoplast using the NtPR1a signal peptide. All used and created plasmids are summarized in **Table S5**.

General level 0 module cloning

pJK006 (pL0M-C1-Rcr3^{pim}) was generated by combining pJK102, pJK103, pJK104, pJK105, pJK106, and pJK107 with pL0V-C1 (EC41308; ENSA consortium) in a BpiI Golden Gate reaction. pJK007 (pL0M-C1-Pip1) was generated by combining pJK110, pJK111, pJK112, pJK113, pJK114, pJK115, and pJK116 with pL0V-C1 (EC41308; ENSA consortium) in a BpiI Golden Gate reaction. pJK253 (pL0M-S-OmpA) was generated by combining the oligo-annealed product of the forward 5'-aatgaaaaagacagctatcgcgattgcagtggcactggctggtttcgctaccgttgcgcaagc-3' and reverse 5'-acctgcttgcgcaacggtagcgaaaccagccagtgccactgcaatcgcgatagctgtcttttt-3' primers with pICH41258 (Addgene #47987 (Weber *et al.*, 2011)) in a BpiI Golden Gate reaction. pJK255 (pL0M-C1-EPIC1) containing EPIC1 lacking the nucleotides encoding for its signal peptide was generated by combining a PCR product generated using pFLAG-EPIC2B (Tian *et al.*, 2007) as a template with the forward 5'-ttgaagacaaaggtcaactgaacggatactcaaag-3' and reverse 5'-ttgaagacaacaccgttggcgggcgtaatcg-3' primers with pL0V-C1 (EC15457, ENSA consortium) in a BpiI Golden Gate reaction.

Binary vector cloning

The binary vector for expression of NbRcr3 containing the residues required for *Cf-2*-dependent recognition of Avr2, pJK479 (pL2M-P19-2x35S::NbRCR3$_{res}$(DPS)) and pJK484 (pL2M-P19-2x35S::NbRCR3$_{res}$(G194N/DPS)), were made by PCR on pJK432 (pL2M-P19-

2x35S::NbRCR3$_{res}$) or pJK483 (pL2M-P19-2x35S::NbRCR3$_{res}$(G194N)) (Chapter 3, this thesis) as a template, respectively, using the forward 5'-ttgaagacaatccttcaggtcattgtgacatcaccaa-3' and reverse 5'-ttgaagacaaaggatctccagtacccttataattctc-3' primers and were circularized in a BpiI Golden Gate reaction. The binary vectors for expression of SmeRcr3 with fragments of Rcr3^{pim}, pJK595-07 (pL2M-P19-2x35S::SmeRCR3-Rcr3^{pim}-Part-1), pJK595-09 (pL2M-P19-2x35S::SmeRCR3-Rcr3^{pim}-Part-3), pJK595-10 (pL2M-P19-2x35S::SmeRCR3-Rcr3^{pim}-Part-4), pJK595-11 (pL2M-P19-2x35S::SmeRCR3-Rcr3^{pim}-Part-5), pJK595-12 (pL2M-P19-2x35S::SmeRCR3-Rcr3^{pim}-Part-6) were made by combining a PCR fragment generated on pJK419-03 (pL2M-P19-2x35S::SmeRCR3) (chapter 3, this thesis) as a template using either the forward 5'- ttgaagacaatctgcagttggagcactaga-3' and reverse 5'-ttgaagacaacggcatgtcatcatcacttatatcatc-3', the forward 5'- ttgaagacaactatccatacaatgggcaacaatac-3' and reverse 5'- ttgaagacaagcagcccatgttgtttgtgg-3', the forward 5'-ttgaagacaacagccagtgtcgattggaattg-3' and reverse 5'-ttgaagacaaatagtctgattcctttgaaattccg-3', the forward 5'-ttgaagacaattggttgttgaagaattcatgg-3' and reverse 5'-ttgaagacaagctgtttggttacggcttgta-3', the forward 5'-ttgaagacaagcttctctagctagagtcgatc-3' and reverse 5'-ttgaagacaaccaatatttctggccctgcc-3' primers, respectively, with either pJK103, pJK105, pJK106, pJK107, or pJK186, respectively in a BpiI Golden Gate reaction. All used and created plasmids are summarized in **Table S5**.

Avr2, Cip1, EPIC1, and EPIC2B purification

The expression vector for periplasmic secretion of N-terminal 6xHis-tagged Cip1, pJK159 (pET28b-T7::OmpA-HIS-TEV-Cip1), was generated by combining pFGH026 (pL0M-C-Cip1) (Grosse-Holz *et al.*, 2018) with pJK082 (pL1V-pET28b), pJK120 (pL0M-S-OmpA-His6-TEV), pJP001 (pL0M-PU-T7-LacO-T7), and pJP002 (pL0M-T-T7 terminator) (chapter 3 this thesis) in a BsaI Golden Gate reaction. The expression vector for periplasmic secretion of C-terminal 6xHis-tagged EPIC1 and EPIC2B, pJK254 (pET28b-T7::OmpA-EPIC1-His)

and pJK256 (pET28b-T7::OmpA-EPIC2B-His), were generated by combining pJK082 (pL1V-pET28b), pJP001 (pL0M-PU-T7-LacO-T7), pJP002 (pL0M-T-T7 terminator) (chapter 3 this thesis), pJK253 (pL0M-S-OmpA), pL0M-C2-6xHis (EC15259, ENSA consortium), and either pJK255 (pL0M-C1-EPIC1) or pFGH084 (pL0M-C1-EPIC2B) in a BsaI Golden Gate reaction. Protein purification was done as described in chapter 3 of this thesis.

Agroinfiltration and HR assays in N. benthamiana

Agroinfiltration and HR assays in *N. benthamiana* were done as described in chapter 3 of this thesis.

HR assays in tomato

Apoplastic fluid (AF) was isolated from agroinfiltrated *N. benthamiana* plants expressing different proteins at 5 days post-infiltration as described in chapter 3 of this thesis. 1 µM Avr2 (final concentration) was added to this AF, which was infiltrated in tomato MM-*Cf-2 rcr3-3* lines (Dixon *et al.*, 2000). Leaves were imaged 2-5 dpi.

ABPP

ABPP was done as described in chapter 3 of this thesis.

P. pastoris vector cloning and protein purification

The *Pichia pastoris* PichiaPink™ yeast expression system was used to produce C-terminally 6xHis-tagged Rcr3^{pim}, Pip1, and Rcr3^{pim}(C154/155A). The high-copy vector pPINKα-HC (ThermoFisher) for use with the PichiaPink™ yeast expression system was made compatible with level 1 Golden Gate assembly using BsaI by removal of the BsaI and BpiI restriction sites. A LacZ selection module was placed between the nucleotides encoding for the dibasic KR Kex2 protease cleavage site of the N-terminal secretion signal from the *S. cerevisiae* α-factor and the *S. cerevisiae CYC1* terminator. pJK081 (pL1VE-SC-pPINKα-HC), for cloning of a

gene from ATG start site (AATG overhang) to stop-codon (GCTT overhang), and pJK101 (pL1VE-C-pPINKα-HC), for cloning of a gene lacking its endogenous signal peptide (AGGT overhang) to stop-codon (GCTT overhang), were made by combining the PCR fragments derived from pPINKα-HC (ThermoFisher) using the forward 5'-ttgaagacaagcttgccggccatttaaatacagg-3' and reverse 5'-ttgaagacaatgacagtgttttcagctaatatttgagc-3', forward 5'-ttgaagacaagtcaagactttcccaggcg-3' and reverse 5'-ttgaagacaatccattgcattcatatacttgg-3', forward 5'-ttgaagacaatggaaaccgaagtgttggg-3' and reverse 5'-ttgaagacaagtcttaatgttccactgacttctgac-3', forward 5'-ttgaagacaaagaccaaacaagactcgcttcg-3' and reverse 5'-ttgaagacaagtcccgcggtatcattgcag-3' primers, and either the forward 5'-ttgaagacaaggacccacgctcaccggctc-3' and reverse 5'- ttgaagacaacattcttttctcgagagataccccttc-3', or the forward 5'-ttgaagacaaggacccacgctcaccggctc-3' and reverse 5'-ttgaagacaaacctcttttctcgagagataccccttc-3' primers with a PCR fragment derived from pICH41308 (Addgene plasmid #47998 (Weber *et al.*, 2011)) using either the forward 5'-ttgaagacaaaatgtgagaccaagcagctggcacgacaggtttcc-3' and reverse 5'-ttgaagacaaaagctgagaccgtcacagcttgtctgtaagcg-3', or the forward 5'-ttgaagacaaaggttgagaccaagcagctggcacgacaggtttcc-3' and reverse 5'-ttgaagacaaaagctgagaccgtcacagcttgtctgtaagcg-3' primers in a BpiI Golden Gate reaction. The methanol-induced *AOX1* promoter drives expression in both constructs. pJK101 (pL1VE-C-pPINKa-HC) was combined with pL0M-C2-6xHis (EC15259; ENSA consortium) and either pJK006 (pL0M-C1-Rcr3^{pim}), pJK149 (pL0M-C1-Rcr3^{pim}(C154/155A)), or pJK007 (pL0M-C1-Pip1) in a BsaI Golden Gate reaction to generate the pJK206 (pPINKα-HC-Rcr3^{pim}-His), pJK209 (pPINKα-HC-Rcr3^{pim}(C154/155A)-His), or pJK211 (pPINKα-HC-Pip1-His), respectively.

10 µg of plasmid DNA of either pJK206, pJK209, or pJK211 was digested with SpeI, heat inactivated, and the digest was ethanol precipitated using 1/10 volume of 3 M sodium acetate

(pH 5.0) and 2.5 volumes of 100% ethanol. DNA was pelleted by centrifugation, washed once with 80% ethanol, air-dried, and resuspended in 10 µl sterile MilliQ. Electrocompetent PichiaPink™ strain 4 (ThermoFisher; *P. pastoris* genotype *ade2, pep4, prb1*) were prepared by inoculating a single colony streaked from a glycerol stock on YPD agar (ThermoFisher) plates grown at 28°C for 5 days in 10 ml of YPD media (ThermoFisher) in a sterile 125 ml baffled flask. This starter culture was incubated at 30°C, 300 rpm for 2 days, and subsequently used to inoculate 100 ml of YPD media in a sterile 1 l baffled flask to an OD_{600} of 0.2. The culture was grown for 2 days at 24°C until the OD_{600} reached 1.3-1.5 (log phase growth). Cells were centrifuged at 1,500 x g at 4°C for 5 min and the pellet was resuspended in 250 ml of ice-cold, sterile MilliQ. Cells were centrifuged as in the previous step, and the pellet was resuspended in 50 ml of ice-cold MilliQ. Cells were centrifuged as in the previous step and the pellet was resuspended in 10 ml of ice-cold 1 M sorbitol. Finally, the cells were centrifuged as in the previous step and the pellet was resuspended in 300 µl of ice-cold 1 M sorbitol for a final volume of approximately 500 µl. Cells were kept on ice prior to use and were used on the same day. Cells were transformed by adding 10 µg of linearized plasmid DNA to 80 µl of the electrocompetent PichiaPink™ cells, and transferred to an ice-cold 0.2 cm electroporation cuvette. The cuvette with the cells was incubated on ice for 5min and pulsed in the electroporator. Immediately after pulsing the cells 1ml of ice-cold YPDS media (ThermoFisher) was added to the cuvette and mixed by pipetting up and down. Cells were incubated at 28°C for at least 2 hours without shaking and after incubation 300 µl of the cell mixtures was plated on PAD selection plates (ThermoFisher) and incubated at 28°C for 3–10 days until distinct colonies formed. White colonies with multiple integrations at the *TRP2* locus were restreaked on fresh PAD selection plates. Glycerol stocks were prepared by culturing a single colony of each strain in 10 ml of YPD medium (ThermoFisher) for 20 hours at 30°C, shaking at 300 rpm. This starter culture was diluted in 200 ml of YPD medium to an OD_{600} of

0.2 and grown for 2 days at 30°C to an OD600 of 2–3. Cells were harvested by centrifugation at 1,500 x g 4°C for 5 min. The pellet was resuspended in YPD medium containing 25% glycerol to a final OD_{600} of 50–100 (approximately 2.5–5.0 × 10^9 cells/ml) and aliquots were snap frozen in liquid nitrogen and store at -80°C.

For protein expression, a single glycerol stock was thawed and used to inoculate a 25 ml starter culture of BMGY (1% yeast extract; 2% peptone; 100 mM potassium phosphate, pH 6.0; 1.34% yeast nitrogen base (YNB) with ammonium sulphate and without amino acids; 0.00004% biotin; 1% glycerol) in a 250 ml baffled flask, grown at 30°C, 300 rpm until culture reached an OD_{600} of 2–6 (approximately 2–3 days). This starter culture was used to inoculate 1 l of BMGY divided over four 1 l baffled flasks and grown at 30°C, 300 rpm until the culture reached log phase growth (OD_{600} = 2–6). Cells were harvested by centrifugation in sterile centrifuge bottles at 3,000 x g for 5 min at room temperature. To induce expression, the supernatant was decanted, and the pellet was resuspended in 200 ml of BMMY medium (1% yeast extract; 2% peptone; 100 mM potassium phosphate, pH 6.0; 1.34% yeast nitrogen base (YNB) with ammonium sulphate and without amino acids; 0.00004% biotin; 0.5% methanol). This culture was divided between two 1 l baffled flasks which were covered with two layers of sterile gauze, and grown for 24 hours at 30°C, 300 rpm. Afterwards the supernatant was collected by centrifugation of the culture at 10,000 x g, 4°C for 20 min. The supernatant was adjusted to 150 mM NaCl, 30 mM imidazole and run over a Ni-NTA agarose (Qiagen)-loaded column by gravity-flow. The Ni-NTA agarose was washed with 10 column volumes of washing buffer (50 mM Tris-HCl, 150 mM NaCl, 50 mM imidazol, pH 8) and the purified protein was eluted in elution buffer (50 mM Tris-HCl, 150 mM NaCl, 250 mM imidazol, pH8). Finally, the protein was concentrated by centrifugation in 3 kDa cut-off filters (Amicon Ultra-15 Centrifugal Filter Units, Merck) at 4,500 x g, 4°C, flash-frozen in liquid nitrogen and stored at -80°C. Protein concentration was determined by Bradford assay using BSA as a standard, and

purity was verified by running the protein on 15% SDS-PAGE gels followed by Coomassie staining (Pink *et al.*, 2010).

PICS peptide library preparation

A modified PICS protocol was used (Biniossek *et al.*, 2016). A PICS peptide library was made by growing a 5 ml *E. coli* DH10B starter culture in LB at 37°C overnight. This starter culture was diluted 1/100 in 400 ml LB, and grown at 37°C to OD_{600} 0.8. Cells were pelleted by centrifugation at 4000 x g at 4°C for 10 min, and the pellet was washed with ice-cold PBS (1.37 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 18 mM KH_2PO_4, pH 7.4). After centrifugation at 4000 x g at 4°C for 10 min the pellet was dissolved in 5 volumes lysis buffer (100 mM HEPES pH 7.5, 150 mM NaCl, 6 M Guanidine-HCl, 10 mM EDTA, 1 mM PMSF (added right before use)) and the cells were lysed by 10 cycles of 10 sec sonication followed by 20 sec resting on ice. The lysed cells were centrifuged for 20 min at 20,000 x g 4°C. The supernatant was collected, and 10 mM dithiothreitol (DTT; final concentration) was added and incubated for 60 min at 25°C. Subsequently 40 mM iodoacetamide (final concentration) was added and incubate for 60 min at 25°C in the dark. The unreacted iodoacetamide was quenched by adding another 10 mM DTT (final concentration 20 mM) followed by a 15 min incubation at 25°C. Proteins were precipitated by chloroform/MeOH precipitation. Briefly, four sample volumes of ice-cold methanol were added to the sample, after vortexing sample was centrifuged at 4,500 x g 10 min. Two sample volume ice-cold chloroform was added, after vortexing sample was centrifuged at 4,500 x g 10 min. Three sample volumes of ice-cold MilliQ was added, after vortexing sample was centrifuged at 4,500 x g 10 min. Upper phase was removed, three sample volumes of ice-cold methanol was added, after vortexing sample was centrifuged at 4,500 x g 10 min. Supernatant was removed and the protein pellet was washed twice with two sample volumes of ice-cold methanol. The pellet was air-dried, and overlayed with ice-cold 20 mM NaOH to an estimated protein concentration of 2.0 mg/ml. 100 mM HEPES pH 7.5 (final

concentration) was added, and centrifuged at 20,000 x g for 10 min at 4°C. The supernatant was collected and digested at 37°C overnight with Trypsin/Lys-C (Promega) at a protease-to-proteome ratio of 1:100 (wt/wt). A small aliquot was analysed by SDS-PAGE to assure complete digestion. The sample was incubated for 20 min at 70°C and 1mM PMSF (final concentration) was added. Peptides were acidified to 0.5% (vol/vol) TFA and degassed by application of a mild vacuum. Peptides were purified on Sep-Pak C18 solid-phase extraction cartridges (Waters) according to the manufactures protocol and eluted in 80% (vol/vol) acetonitrile without added TFA. Acetonitrile was removed by vacuum evaporation to a final concentration of approximately 2.5% (vol/vol), and after centrifugation at 20,000 x g, 4°C for 10 min the supernatant was collected and stored at a peptide concentration of 2 mg/ml at -80°C.

PICS digestion

100 µg of peptide library was used per treatment. The peptide library was diluted to 1 mg/ml (final concentration) in 50 mM NaAc pH 5 and 10 mM tris(2-carboxyethyl)phosphine (TCEP; final concentration). Purified protease ($Rcr3^{lyc}$ or Pip1) or control (buffer) were matured in 50 mM NaAc pH 5 and 10 mM TCEP (final concentration) for 5 hours, prior to addition at a ratio of 1:100 (wt/wt) and incubation for 1hr at room temperature. 100 µM E-64 (final concentration) was added to inactivate the proteases, and proteases were further heat-inactivated for 5 min at 95°C. Protease-treated and control samples were isotopically labelled by reductive dimethylation. $Rcr3^{lyc}$-, Pip1-, and buffer-treated samples were labelled with either 20 mM $^{12}CH_2O$ (cat. no. 252549, Sigma-Aldrich) and 20 mM $NaBH_3CN$ (cat. no. 252549, Sigma-Aldrich) (light), 20 mM $^{12}CD_2O$ (Sigma-Aldrich) and 20 mM $NaBH_3CN$ (medium), or 20 mM $^{13}CD_2O$ (cat. no. 596388, Sigma-Aldrich) and 20 mM $NaBD_3CN$ (Sigma-Aldrich) (heavy), respectively, and incubated overnight at 37°C. Afterwards an additional 20 mM formaldehyde (final concentration 40 mM) and 20 mM sodium cyanoborohydride (40 mM final concentration) was added to ensure complete labelling. Labelling was quenched by the addition

of 100 mM Tris-HCl pH 6.8 and 60 min incubation at 37°C. Following labelling, protease-treated and control samples were mixed at equal ratios and desalted by C18 solid phase extraction (Sep-Pak, Waters) according to manufacturer's instructions and concentrated to 10 µl using a vacuum concentrator (ThermoFisher).

PICS MS/MS analysis

3 µl of the desalted samples were loaded onto a C18 reverse phase capillary trap column (Acclaim PepMap C18, 75 µm x 2 cm, 3 µm particle size, ThermoFisher) and separated using a C18 reverse phase analytical column (Acclaim PepMap C18, 75 µm x 25 cm, 2 µm particle size, 100 Å, ThermoFisher) coupled to an UltiMate3000 nano RSLC system (Thermo). A 90 min gradient from 5 to 32% (v/v) acetonitrile 0.1% formic acid in H_2O was used to elute the peptides at a flow rate of 300 nl min^{-1}. The nano-LC system was on-line coupled to an Impact II high-resolution quadrupole-time of flight tandem mass spectrometer using a CaptiveSpray nano-electrospray source (Beck *et al.*, 2015). MS spectra were acquired in a range from m/z 200 to 1750 at 4 Hz. For fragmentation, the 17 most intense precursor ions of each MS scan were selected (Top 17 method). MS/MS spectra of the fragmented precursor ions were recorded in a mass range from m/z 300 to 1750 at an intensity-dependent collection rate of 2 to 20 Hz spectrum.

Peptides were identified and quantified using the MaxQuant software package, version 1.6.0.16 (Tyanova *et al.*, 2016) using the Andromeda search engine (Cox *et al.*, 2011). Generic settings for Bruker Q-TOF instruments were used to match spectra to protein sequences to the *E. coli* proteome database (strain K12, reference proteome, 4308 entries). Search parameters included precursor mass tolerance of ±10 ppm, fragment ion mass tolerance of ±20 ppm, semi-tryptic peptides with up to one missed cleavage, static modifications were cysteine carboxyamidomethylation (+57.02 Da), lysine and N-terminal dimethylation (light

formaldehyde 28.0313 Da; medium formaldehyde 32.0564 Da; heavy formaldehyde 36.0756 Da) and methionine oxidation and deamidation of asparagine and glutamine as variable modifications. A false discovery rate of 0.01 was applied both for spectrum-to-sequence matching and protein identification. A custom PERL script was used to filter semi-specific peptides and through database-lookup determine the corresponding prime and nonprime sequence.

Acknowledgements.

We thank Ursula Pyzio for excellent plant care; John Baker for photography; Sylvestre Marillonnet and Nicola Patron for providing plasmids via AddGene.

Funding. This work has been supported by 'The Clarendon Fund' (JK), and the ERC Consolidator grant 616449 'GreenProteases' (RvdH, JK). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests. The authors have declared that no competing interests exist.

Author contributions. Conceptualization: JK, RvdH; Formal analysis: JK; Investigation: JK, and OM, SK, ROG, SR and PSK under supervision of JK; PICS analysis: FD, PH; Funding acquisition: RvdH; Writing: JK, RvdH.

Supplemental figures

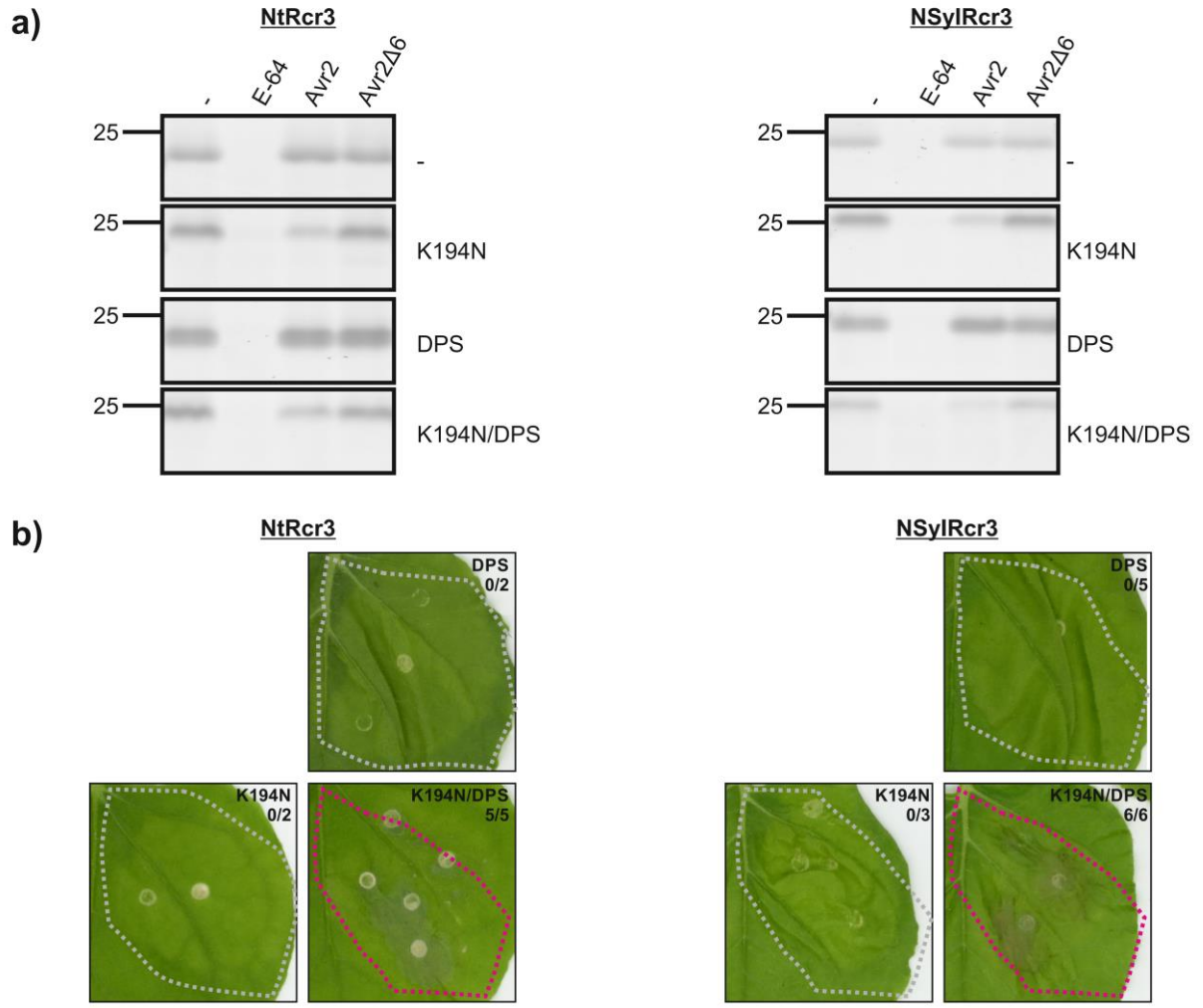


Figure S1: *N. sylvestris* and *N. tabacum* Rcr3 can be engineered to trigger a Cf-2/Avr2-dependent HR. **a)** Activity-based protein-profiling of Papain-like cysteine proteases. Rcr3^{pim}, and *N. sylvestris* and *N. tabacum* Rcr3 mutants (NSylRcr3 and NtabRcr3, respectively), were co-expressed with silencing inhibitor P19 by agroinfiltration of *N. benthamiana* leaves. Apoplastic fluids were isolated at 6 days-post-infiltration (dpi), pre-incubated for 45 min with and without 100 µM E-64, 2 µM Avr2 or Avr2Δ6, and labelled for 3 hours with 2 µM MV201. Samples were separated on SDS-PAGE gels and scanned for fluorescence. **b)** Co-infiltration of *A. tumefaciens* containing binary vectors for the expression of Cf-2, Avr2, and either NSylRcr3 or NtabRcr3 or the respective K194N, DPS, and K194N/DPS mutants. Only the K194N/DPS mutants of NSylRcr3 and NtabRcr3 trigger a HR in *N. benthamiana*. *A. tumefaciens* containing different binary vectors were combined at a final OD_{600} = 0.25 each in a 1:1:1 ratio. The silencing inhibitor P19 was present in the backbone of all binary vectors. Pictures were taken at 5 dpi. Numbers indicate the number of leaves showing HR.

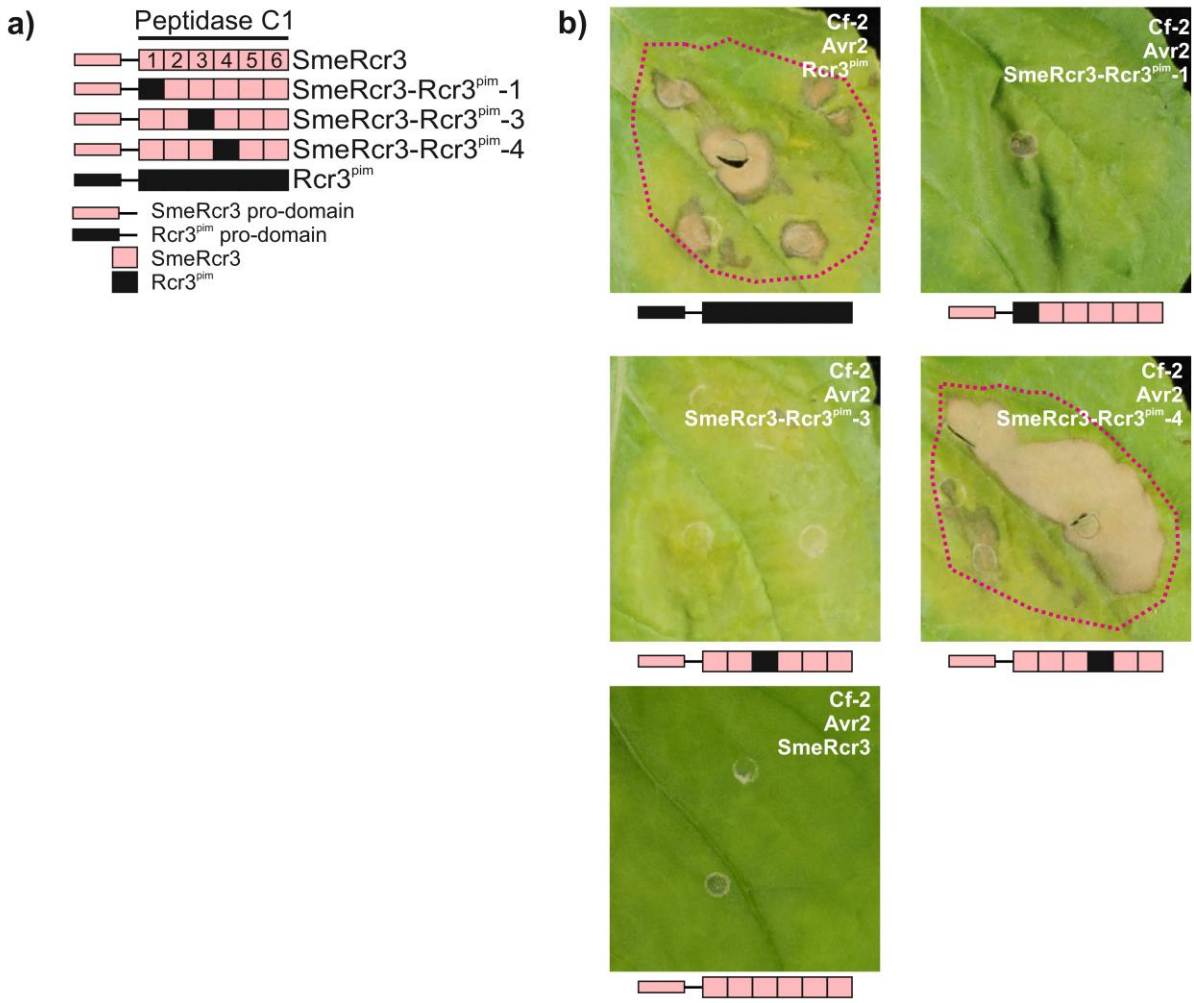


Figure S2: Fragment-swapping between tomato Rcr3 and Eggplant Rcr3 identifies a region in Rcr3 required for triggering a Cf-2/Avr2-dependent HR. **a)** Design of fragment-swapped eggplant Rcr3 (SmeRcr3) containing fragments of $Rcr3^{pim}$. **b)** Only the transient expression of SmeRcr3 with fragment 4 of $Rcr3^{pim}$ with Cf-2 and Avr2 triggers a HR in *N. benthamiana*. *A. tumefaciens* containing different binary vectors were combined at a final OD_{600} = 0.25 each in a 1:1:1 ratio. The silencing inhibitor P19 was present in the backbone of all binary vectors. Pictures were taken at 5 dpi.

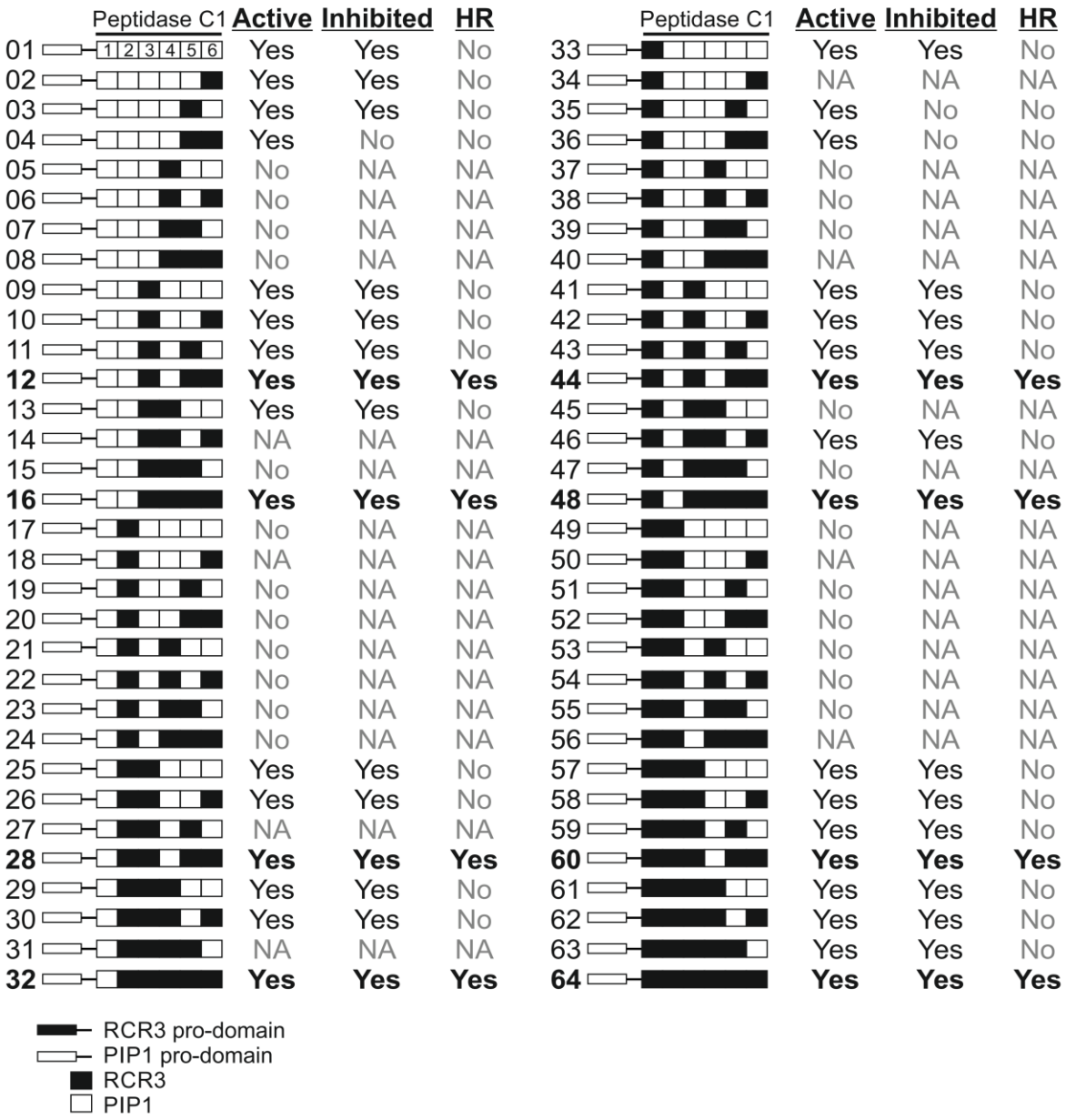


Figure S3: Design of fragment-swapped Rcr3/Pip1 hybrids. The peptidase C1 domain (PF00112)) Rcr3 and Pip1 were split up into six fragments. Out of 64 possible combinations with the Pip1 pro-domain, 57 constructs were made. 33 of these were active, and 30 could be inhibited by Avr2. Eight could trigger *Cf-2*-dependent Avr2 recognition, all of which contained Rcr3 fragment 3, 5, and 6. **Table S1** shows the Rcr3/Pip1 variant residues of each fragment.

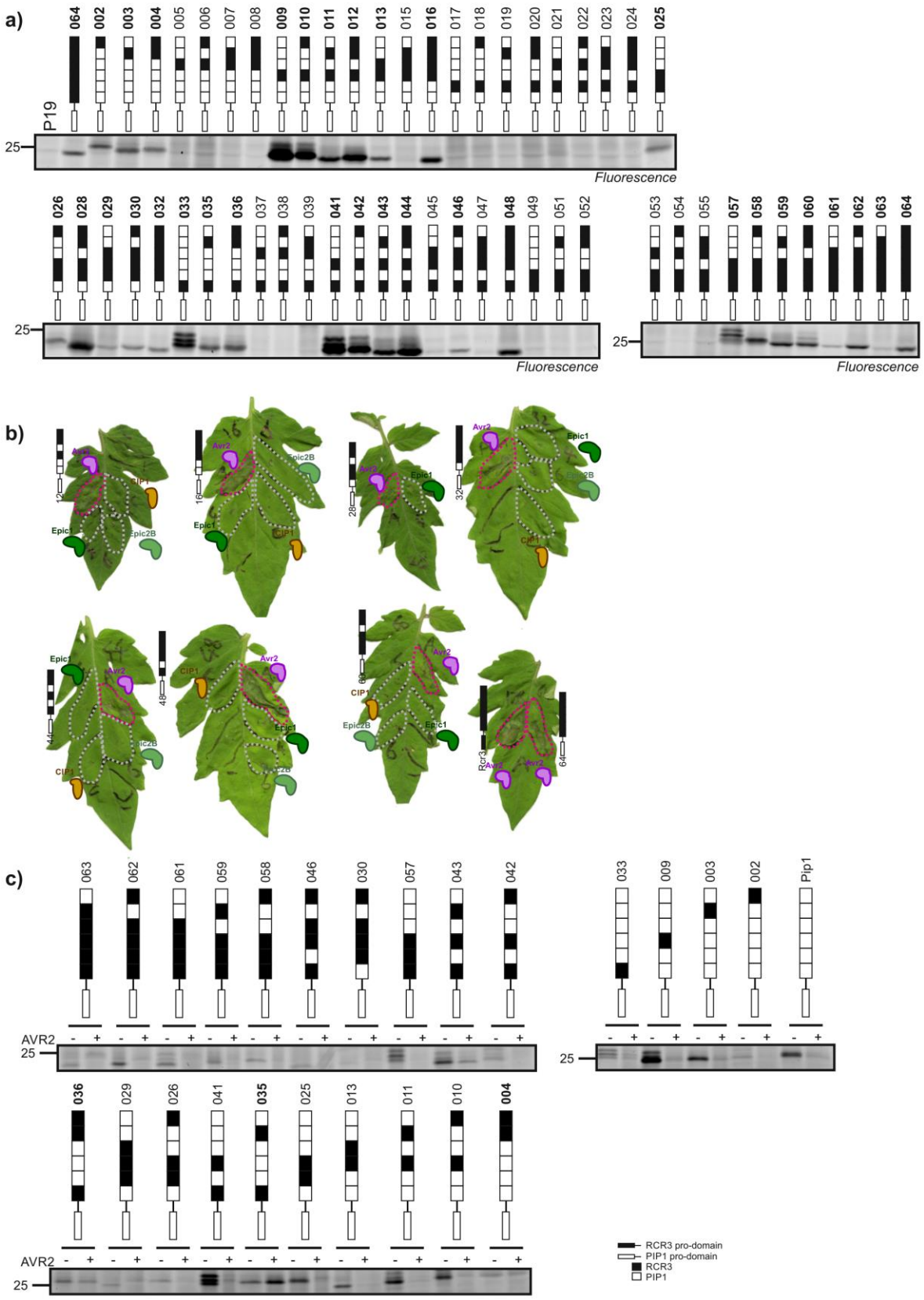


Figure S4: Rcr3/Pip1 fragment-swapped hybrids are active proteases which can be inhibited by Avr2 and trigger a specific Cf-2/Avr2-dependent HR. **a)** Activity-based protein-profiling of Papain-like cysteine proteases. Rcr3/Pip1 hybrids were co-expressed with silencing inhibitor P19 by agroinfiltration of *N. benthamiana* leaves. Apoplastic fluids were

Figure S4. (Continued). isolated at 5 days-post-infiltration (dpi) and labelled for 5 hours with 2µM MV201. Samples were separated on SDS-PAGE gels and scanned for fluorescence. Active hybrid proteases are indicated in bold. **b)** Co-injection of apoplastic fluids containing Rcr3/Pip1 fragment-swapped hybrids produced as above with either 1µM *E. coli* produced recombinant Avr2, Cip1, EPIC1, or EPIC2B in tomato MM-*Cf-2 rcr3-3* plants. Only the co-injection with Avr2 triggers *Cf-2*-dependent HR. Pictures were taken at 5dpi. **c)** Activity-based protein-profiling of Papain-like cysteine proteases. Apoplastic fluids containing Rcr3/Pip1 fragment-swapped hybrids which cannot trigger in Cf-2/Avr2-dependent HR produced as above, were pre-incubated for 45min with and without 500nM Avr2 and labelled for 5 hours with 2µM MV201. Samples were separated on SDS-PAGE gels and scanned for fluorescence. Hybrid proteases which cannot be inhibited by Avr2 are indicated in bold.

Supplemental information

Table S1: Summary of variant residues between Rcr3 and Pip1 in the peptidase C1 domain.

Table S2: Fragment three tested hybrids and results.

Table S3: Fragment five tested hybrids and results

Table S4: Fragment six tested hybrids and results

Table S5: Plasmids used in this study.

Supplemental dataset 1: Complete gels of figures.

Chapter 6

Re-annotated *Nicotiana benthamiana* gene models for enhanced proteomics and reverse genetics

Kourelis, J., Kaschani, F., Grosse-Holz, F.M., Homma, F., Kaiser, M. and van der Hoorn, R.A.L. (2018) Re-annotated *Nicotiana benthamiana* gene models for enhanced proteomics and reverse genetics. *bioRxiv*, 373506.

In preparation for re-submission

Authors

Jiorgos Kourelis[1], Farnusch Kaschani[2], Friederike M. Grosse-Holz[1], Felix Homma[1], Markus Kaiser[2], Renier A. L. van der Hoorn[1]

Affiliations

[1]Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, South Parks Road, OX1 3RB Oxford, UK

[2]Chemische Biologie, Zentrum fur Medizinische Biotechnologie, Fakultät für Biologie, Universität Duisburg-Essen, Essen, Germany.

Author contributions

Conceptualization: JK, RvdH; Formal analysis: JK; Funding acquisition: RvdH; Wetlab experiments: FHG; Proteomics: FK, MK; Programming: JK, FH; Writing: JK, RvdH.

Justification

For this paper I re-annotated the *N. benthamiana* gene models and used the resulting database to compare the extracellular (apoplastic) proteome of leaves with a total extract of leaves. Additionally I investigated the subtilisin gene family in *N. benthamiana*. This research takes a broader approach than my previous chapters, looking at which enzymes are present in the apoplast and may have a role in immunity there. Additionally, I extend the concept of pseudogenization in *N. benthamiana*, as I showed for the PLCP gene family in chapter 3.

Re-annotated *Nicotiana benthamiana* gene models for enhanced proteomics and reverse genetics

Jiorgos Kourelis[1], Farnusch Kaschani[2], Friederike M. Grosse-Holz[1], Felix Homma[1], Markus Kaiser[2], Renier A. L. van der Hoorn[1]

[1]Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, South Parks Road, OX1 3RB Oxford, UK; [2]Chemische Biologie, Zentrum fur Medizinische Biotechnologie, Fakultät für Biologie, Universität Duisburg-Essen, Essen, Germany.

***Nicotiana benthamiana* is an important model organism and representative of the Solanaceae (Nightshade) family. *N. benthamiana* has a complex ancient allopolyploid genome with 19 chromosomes, and an estimated genome size of 3.1Gb. Several draft assemblies of the *N. benthamiana* genome have been generated, however, many of the gene-models in these draft assemblies appear incorrect. Here we present a nearly non-redundant database of 42,855 improved *N. benthamiana* gene-models. With an estimated 97.6% completeness, the new predicted proteome is more complete than the previous proteomes. We show that the database is more sensitive and accurate in proteomics applications, while maintaining a reasonable low gene number. As a proof-of-concept we use this proteome to compare the leaf extracellular (apoplastic) proteome to a total extract of leaves. Several gene families are more abundant in the apoplast. For one of these apoplastic protein families, the subtilases, we present a phylogenetic analysis illustrating the utility of this database. Besides proteome annotation, this database will aid the research community with improved target gene selection for genome editing and off-target prediction for gene silencing.**

Keywords: *Solanaceae // Genome annotation // Nicotiana benthamiana // Proteomics // Subtilases*

Database availability. The database has been uploaded at Oxford Research Archives (ORA) and can be downloaded from this link: https://ora.ox.ac.uk/objects/uuid:f09e1d98-f0f1-4560-aed4-a5147bc7739d. We hope that this database will be accessible via additional databases in the near future.

Introduction

Nicotiana benthamiana has risen to prominence as a model organism for several reasons. First, *N. benthamiana* is highly susceptible to viruses, resulting in highly efficient virus-induced gene-silencing (VIGS) for rapid reverse genetic screens (Senthil-Kumar and Mysore, 2014). This hypersusceptibility to viruses is due to an ancient disruptive mutation in the RNA-dependent RNA polymerase 1 gene (*Rdr1*), present in the lineage of *N. benthamiana* which is used in laboratories around the world (Bally *et al.*, 2015). Reverse genetics using *N. benthamiana* have confirmed many genes important for disease resistance (Wu *et al.*, 2017; Senthil-Kumar *et al.*, 2018). Additionally, *N. benthamiana* is highly amenable to the generation of stable transgenic lines (Clemente, 2006; Sparkes *et al.*, 2006) and to transient expression of transgenes (Goodin *et al.*, 2008). This easy manipulation has facilitated rapid forward genetic screens and has established *N. benthamiana* as the plant bioreactor of choice for the production of biopharmaceuticals (Stoger *et al.*, 2014). Finally, *N. benthamiana* is a member of the Solanaceae (Nightshade) family which includes important crops such as potato (*Solanum tuberosum*), tomato (*Solanum lycopersicum*), eggplant (*Solanum melongena*), and pepper (*Capsicum* ssp.), as well as tobacco (*Nicotiana tabacum*) and petunia (*Petunia* ssp.).

N. benthamiana belongs to the *Suaveolentes* section of the *Nicotiana* genus, and has an ancient allopolyploid origin (>10Mya) accompanied by chromosomal re-arrangements resulting in a complex genome with 19 chromosomes in the haploid genome – reduced from the ancestral allotetraploid 24 chromosomes - and an estimated haploid genome size of ~3.1Gb (Leitch *et al.*, 2008; Goodin *et al.*, 2008; Wang and Bennetzen, 2015). There are four independent draft assemblies of the *N. benthamiana* genome (Bombarely *et al.*, 2012; Naim *et al.*, 2012), as well as a de-novo transcriptome generated from short-read RNAseq (Nakasugi *et al.*, 2014). These datasets have greatly facilitated research in *N. benthamiana*, allowing for efficient prediction of off-targets of VIGS (Fernandez-Pozo *et al.*, 2015) and genome editing using CRISPR/Cas9 (Liu *et al.*, 2017), as well as RNAseq and proteomics studies (Grosse-Holz *et al.*, 2018). These draft assemblies are, however, several years old and gene annotations have not been updated since. Furthermore, in the course of our research we realized that many of the gene models in these draft assemblies are incorrect, and that putative pseudo-genes are often annotated as protein-encoding genes. This is exacerbated because these draft assemblies are highly fragmented and that *N. benthamiana* has a complex origin. Furthermore, the *de-novo* transcriptome assembly has a high proportion of chimeric transcripts. Because of incorrect annotations, extensive processing is required to select target genes for reverse genetic approaches such as gene silencing and editing, or for phylogenetic analysis of gene families. We realized that gene annotation in several other species in the *Nicotiana* genus were much better (Xu *et al.*, 2017; Sierro *et al.*, 2013; Sierro *et al.*, 2014), and decided to re-annotate the available *N. benthamiana* draft genomes using these gene models as a template. The gene models obtained in this way were extracted into a single non-redundant database with improved gene models. Here we show that this database is more accurate and sensitive for proteomics, facilitates phylogenetic analysis of gene families, and may be useful for genome editing and VIGS on- and off-target prediction.

Results and Discussion

Re-annotation of gene-models in the N. benthamiana genome assemblies

For the annotation of gene-models in the different *N. benthamiana* draft genomes, we chose to use Scipio (Keller *et al.*, 2008). Scipio refines the transcription start-site, exon-exon boundaries, and the stop-codon position of protein sequences aligned to the genome using BLAT (Keller *et al.*, 2008). Importantly, given that the input protein sequences are well-annotated, this method is more accurate and sensitive than other gene prediction methods (Keller *et al.*, 2011). Because the efficiency of this process correlates with phylogenetic distance, we took the predicted protein sequences from recently sequenced *Nicotiana* species (**Figure 1**) (Sierro *et al.*, 2013; Xu *et al.*, 2017). We then used CD-HIT at a 95% identity cut-off to reduce the redundancy in this database and additionally to remove partial sequences (**Figure 1**, Step 1). The resulting database – Nicotiana_db95 – contains 85,453 protein sequences from various *Nicotiana* species. We used the protein sequences in this database as an input to annotate gene-models in the four independent draft assemblies of the *N. benthamiana* genome using Scipio (**Figure 1**, Step 2). As the available *N. benthamiana* draft genomes are highly fragmented and each individual draft genome may miss a number of genes, we extracted the gene-models generated by Scipio, filtered for redundancy using CD-HIT-EST and combined the gene-models into a single database (NbA) containing 41,651 gene-models (**Figure 1**, Step 3). We next compared the predicted proteome derived from our NbA database against the published predicted proteomes using a proteomics dataset from a full-leaf extract or apoplastic fluid (samples described further in the manuscript). Proteins for which peptides were identified in the published proteomes but not in our NbA database were extracted and re-annotated in the draft genomes as described above and added to our NbA database resulting in the NbB database containing 42,884 gene-models (**Figure 1**; Step 4, 1233 additional entries). Finally, missing BUSCOs (Benchmarking Universal Single-Copy Orthologues) (Simão *et al.*,

2015; Waterhouse *et al.*, 2018) were re-annotated in our database as above, together with the manual curation of several gene-families in which several duplicated genes were removed (see Material & Methods) to obtain our final NbC database containing 42,853 entries (**Figure 1**; Step 5).

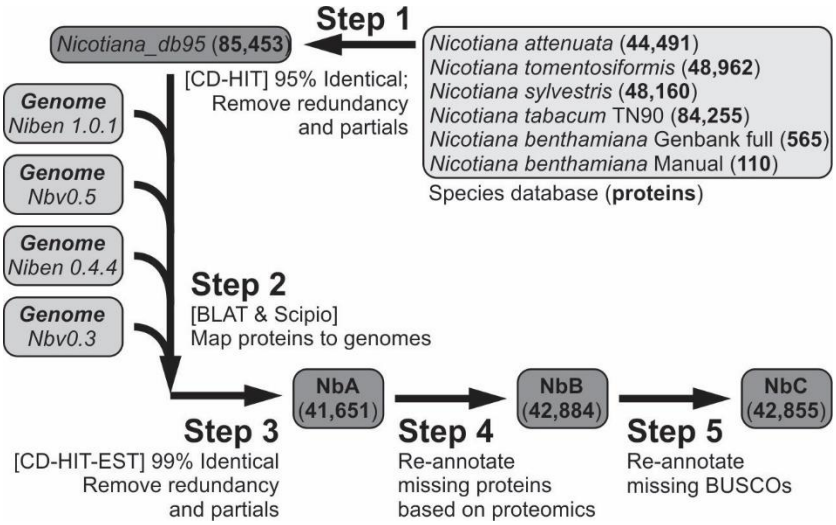


Figure 1. Bioinformatics pipeline for improved *Nicotiana benthamiana* proteome annotation. A database of *Nicotiana* predicted protein sequences was retrieved from Genbank and clustered at 95% identity threshold to reduce redundancy (Step 1), and used to annotate the four available *N. benthamiana* draft genomes (Step 2). The databases derived from the different genome builds were clustered at 99% sequence identity using CD-HIT-EST-2D in the following order: Niben1.0.1 > Nbv0.5 > Niben0.4.4 > Nbv0.3 generating NbA (Step 3). Proteins identified by proteomics in other databases but missing in NbA were added as above to generate NbB (Step 4), and finally missing BUSCOs were added as above to generate the final database NbC (Step 5).

The new proteome database is more complete, more sensitive and accurate, and relatively small

We next compared the predicted proteome database to the published predicted proteomes. We also included our Nicotiana_db95 proteome database in this comparison. The published proteomes included the predicted proteomes from the Niben0.4.4 and Niben1.0.1 draft genomes, a previously described curated database in which gene-models from Niben1.0.1 were corrected using RNAseq reads (Grosse-Holz *et al.*, 2018), and the predicted proteome derived from the *de-novo* transcriptomes (Nakasugi *et al.*, 2014).

We used BUSCO (Simão *et al.*, 2015; Waterhouse *et al.*, 2018) as a quality measure to estimate the completion of our database as compared to the published predicted proteomes (**Figure 2a**). The BUSCO set used contains 1440 highly conserved plant genes which are expected to be predominantly found in a single-copy (Simão *et al.*, 2015). Nicotiana_db95 has one fragmented and nine missing BUSCOs, indicating that at best we should be able to identify 99.3% of the *N. benthamiana* genes using this database (**Figure 2a**). In our NbB database, 1406/1440 (97.6%) BUSCO proteins were identified as complete, of which 762 were single-copy (54.2%), 644 were duplicated (45.8%), nine sequences were fragmented (0.63%), and 25 were missing (1.74%) (**Figure 2a**). The high number of duplication is likely due to either technical duplication generated by small variations between the different draft genomes, or genuinely duplicated genes arising from the allo-tetraploid origin of *N. benthamiana*. By adding missing BUSCOs into the NbC database, we recovered eight of the nine fragmented BUSCOs and ten of the 25 missing BUSCOs. In comparison, the next most complete, previously published proteome is the predicted proteome from the Nbv5.1 primary + alternate transcriptome, which has 12 fragmented and 32 missing BUSCOs, but it also has nearly five times more proteins than our database and 71.8% of BUSCOs are duplicated.

Next, we investigated the number of unique PFAM identifiers found with each entry in each proteome, as an estimation of the number of proteins incorrectly annotated (**Figure 2b**). We expect that miss-annotated sequences and fragmented gene products are less likely to get a PFAM annotation. Indeed, significantly more proteins get at least one PFAM identifier in our three databases as compared to the published proteomes, indicating that proteins in our database are better annotated.

Furthermore, we looked at the length distributions of proteins in the different predicted proteomes (**Figure 2c**). We reasoned that the protein-length distribution should be similar to that of the Nicotiana_db95 database. The proteins in the final proteome are significantly longer

than those in the Niben0.4.4, Niben1.0.1, and manually curated proteome (Grosse-Holz et al., 2018) while the proteins in the Nbv5.1 primary + alternate proteome are on average larger than in our final NbC database. We speculate that the Niben0.4.4 and Niben1.0.1 predicted proteomes contain many pseudo-genes which are annotated as protein-encoding as well as partial genes (**Figure 2c**), while the Nbv5.1 primary + alternative proteome has a high proportion of chimeric sequences which due to the short-read sequencing techniques used are biased towards long transcripts (**Figure 2c**). Additionally, the curated proteome has a large proportion of very small proteins and 47.3% of genes do have a PFAM annotation, which we speculate is due to partial sequences or spurious small ORFs being annotated as protein-encoding (**Figure 2b,c**).

Finally, comparing the different proteomes on a proteomics dataset indicates the new database has the highest sensitivity, with the highest percentage of annotated MS/MS spectra in both tested samples, while it has the fewest entries (**Figure 2d**). Additionally, using our new NbC database, we identify the highest number of unique peptides identified in at least 3 out of 4 biological replicates of both proteomes (**Figure 2d**). These metrics combined indicate that the new NbC database is more sensitive and accurate for proteomics than the currently available databases. Importantly, this does not come at the cost of increased redundancy, which would hinder downstream applications.

Since the previous proteomics dataset was also used to re-annotate gene-models (**Figure 1**; Step 4), we independently validated our database on an independent dataset where we re-analysed a previously published apoplastic proteome of agro-infiltrated *N. benthamiana* as compared to non-infiltrated *N. benthamiana* (PRIDE repository PXD006708) (Grosse-Holz et al., 2018). The new NbC database was also more sensitive and accurate than the Curated database on this dataset (18,430 vs 17,960 peptides detected, 22.5%±/-3.1% vs 21.7%±3.0% spectra identified).

Finally, since phylogenetic analysis of gene families in closely related species often relies on gene-annotations, we compared the predicted proteome from our NbB database against the predicted proteomes of Solanaceae species for which genomes have been sequenced (**Figure S1a,b**). Our NbB proteome compares well to the predicted proteomes of other sequences Solanaceae species. Additionally, since the predicted proteomes of some of these species miss a relatively high proportion of genes (up to 28.5% of genes missing or fragmented), care must be taken to not over-interpret results derived from phylogenetic analysis using these sequences.

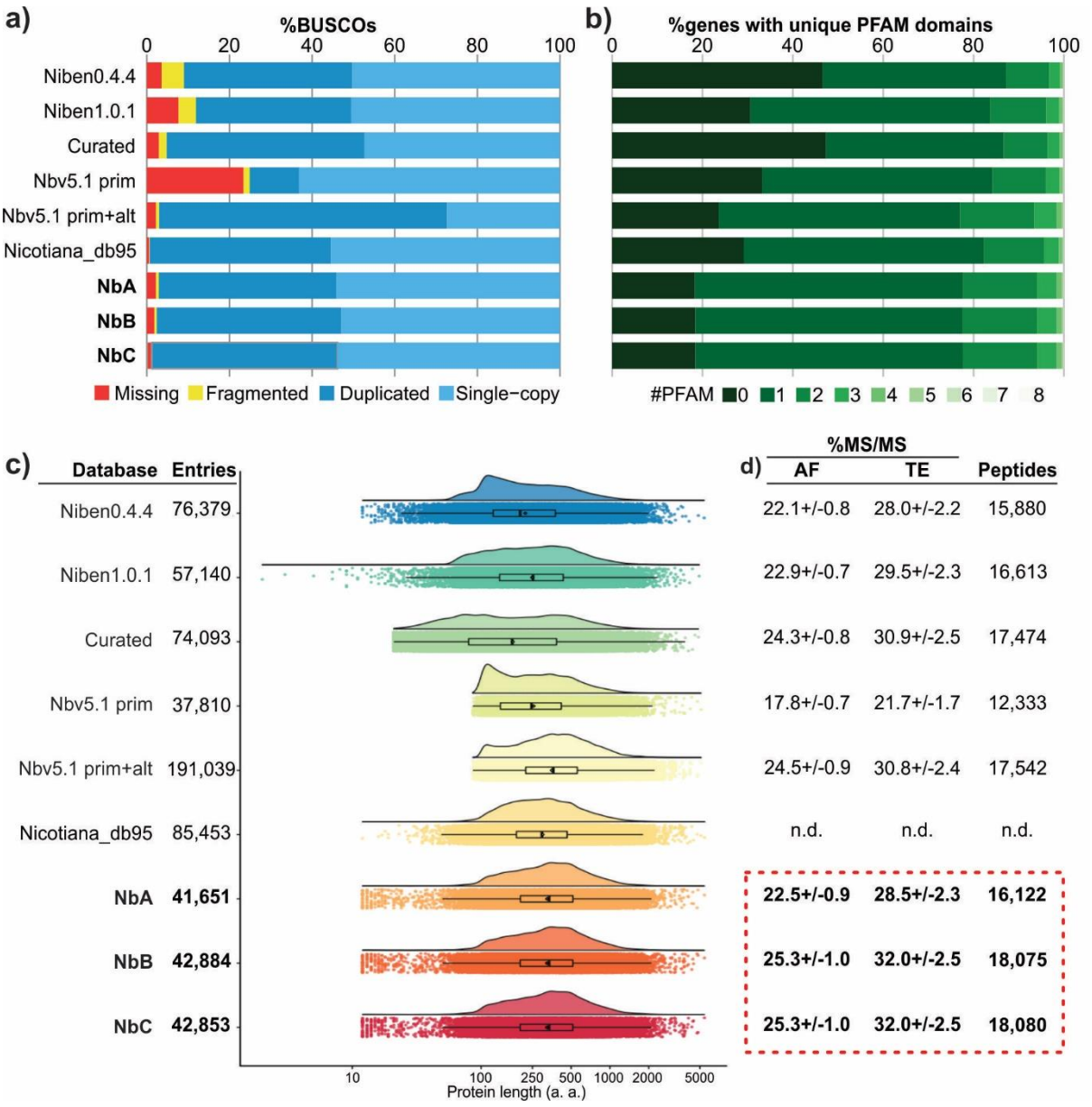


Figure 2. Increased coverage and annotation of *N. benthamiana* proteins.
a) Completeness of the different predicted proteomes was estimated using BUSCO with the embryophyta database. **b)** The percentage of the proteins assigned with a certain number of unique PFAM identifiers (from 0-8) is plotted. **c)** Violin and boxplot graph of $\log_{10}$ protein length distribution of each database. Jittered dots show the raw underlying data. **d)** Percentage of annotated MS/MS spectra in AF or TE samples. Means and standard deviation (n=4) are shown and the number of unique peptides identified in at least of three out of four biological replicates of either AF or TE.

Improved annotation of the apoplastic proteome of N. benthamiana

Next we used our final NbC database to analyse the extracellular protein repertoire of the *N. benthamiana* apoplast. The plant apoplast is the primary interface in plant-pathogen interactions (Misas-Villamil and van der Hoorn, 2008; Doehlemann and Hemetsberger, 2013) and apoplastic proteins include many enzymes potentially important in plant-pathogen interactions. We found the protein composition of leaf apoplastic fluid (AF) to be distinct from that of a leaf total extract (TE) (**Figure 3a**). We considered proteins apoplastic when only detected in the AF samples or those with a $\log_2$ fold abundance difference $\geq$1.5 and a p-value cut-off off $\leq$0.01 (BH-adjusted moderated t-test) in the comparison of AF vs TE (518 proteins). Similarly, we considered proteins intracellular when found only in the TE or those proteins with a $\log_2$ fold abundance difference $\leq$-1.5 with a p-value cut-off off $\leq$0.01 in the comparison of AF vs TE (1042 proteins) (**Figure 3b**). The remainder proteins was considered both apoplastic and intracellular (832 proteins). As expected, the apoplastic proteome is significantly enriched for signal peptide containing proteins, while the intracellular proteins and proteins present both in the apoplast and intracellular are significantly enriched for proteins lacking a signal peptide (BH-adjusted hypergeometric test, $p<0.001$).

Proteins considered predominantly intracellular are enriched for GO-SLIM terms associated with translation, photosynthesis and transport as biological processes (**Figure 3c**), and a similar pattern is seen for the molecular function terms (**Figure 3d**). Proteins present both in TE and in AF are enriched for GO-SLIM terms associated with biosynthetic processes,

and homeostasis (**Figure 3c**). These processes usually performed by proteins acting at multiple subcellular localizations. The apoplastic proteome is enriched for proteins acting in catabolic processes and carbohydrate and lipid metabolic processes (**Figure 3c**), which is reflected in the enrichment of peptidases and glycosidases (**Figure 3d, Table S1** for a full list).

To specify which peptidases are enriched in the apoplast, we also annotated the proteome with MEROPS peptidase identifiers (Rawlings *et al.*, 2018). Three of the 15 different families of peptidases detected in the apoplast have significantly more members enriched in the AF as compared to TE, namely the subtilase (S08; 13 members, $p<0.001$), serine carboxypeptidase-like (S10; 8 members, $p<0.01$), and aspartic peptidase families (A01; 16 members, $p<0.001$), while the proteasome is enriched in the intracellular fraction (T01; 27 members, $p<0.001$) (BH-adjusted hypergeometric test, **Table S2** for a full list).

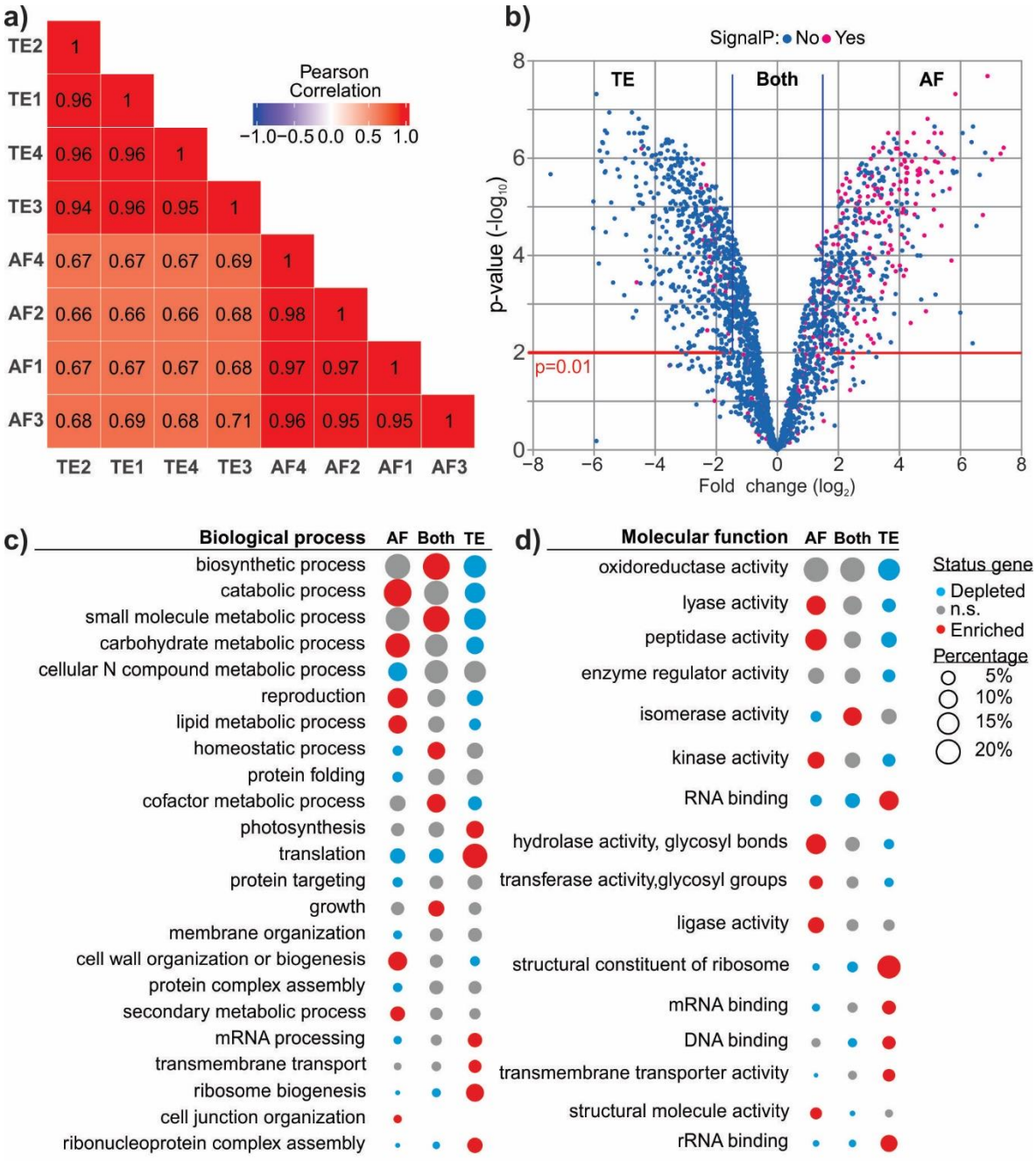


Figure 3. Improved annotation of the *N. benthamiana* apoplastic proteome.
a) Correlation matrix heat map of the LFQ intensity of protein groups in the biological replicates of AF and TE. Biological replicates are clustered on similarity. **b)** A volcano plot is shown plotting $\log_2$ fold difference of AF/TE over $-\log_{10}$ BH-adjusted moderated p-values. Proteins $\log_2$ fold change ≥1.5 and p≤0.01 were considered apoplastic, as well as those only found in AF. Conversely, proteins with a $\log_2$ fold change ≤-1.5 and p≤0.01 were considered strictly intracellular, as well as those found only in TE. **c)** and **d)** shows a grid where each row represents an GO-SLIM annotation significantly enriched or depleted (BH-adjusted hypergeometric test, p<0.05) in at least one of the fractions (apoplast, intracellular, or both) and each column the compartment. Each bubble indicates the percentage of genes containing that specific GO-SLIM annotation in that compartment. Colours indicate whether the GO-SLIM annotations are enriched or depleted in that fraction (p<0.05, n.s. = non-significant).

Figure 3. (Continued). **c)** Percentage of proteins in each fraction annotated with biological process-associated GO-SLIM terms; **d)** Molecular function-associated GO-SLIM terms.

Pseudogenization in the subtilisin family is consistent with a contracting functional genome

One of the gene families found enriched in the apoplast is the subtilisin family. Several subtilisins are implicated in immunity, notably the tomato P69 clade of subtilisins (Taylor and Qiu, 2017). In order to estimate the completeness of our database, we manually verified and corrected genes belonging to the subtilisin gene family. Our NbC database contains 64 complete subtilisin genes, and one partial gene. By searching the Niben1.0.1 and Niben0.4.4 genome assemblies, we identified an additional 43 putative subtilisin pseudo-genes which had internal stop-codons and are therefore likely non-coding.

Interestingly, phylogenetic analysis shows that close paralogs are often pseudogenised. This pattern of pseudogenization in the subtilisin gene family is consistent with a contracting functional genome upon polyploidization, where for each functional protein-encoding gene there is a corresponding pseudo-gene (**Figure 4**, and **Figure S2**). Remarkably, no SBT3 clade family members were identified in *N. benthamiana* (**Figure S3**). Finally, we looked for the amino acid residue at the pro-domain junction, as the presence of an aspartic acid residue is indicative of phytaspase activity (Reichardt et al., 2018). Three *N. benthamiana* subtilisins may possess phytaspase activity based on the presence of an apartic acid residue at the pro-domain junction as well as a histidine residue in the S1 pocket which is thought to bind to P1 aspartic acid (**Figure S2**, and **Figure S3**, Reichardt et al., 2018).

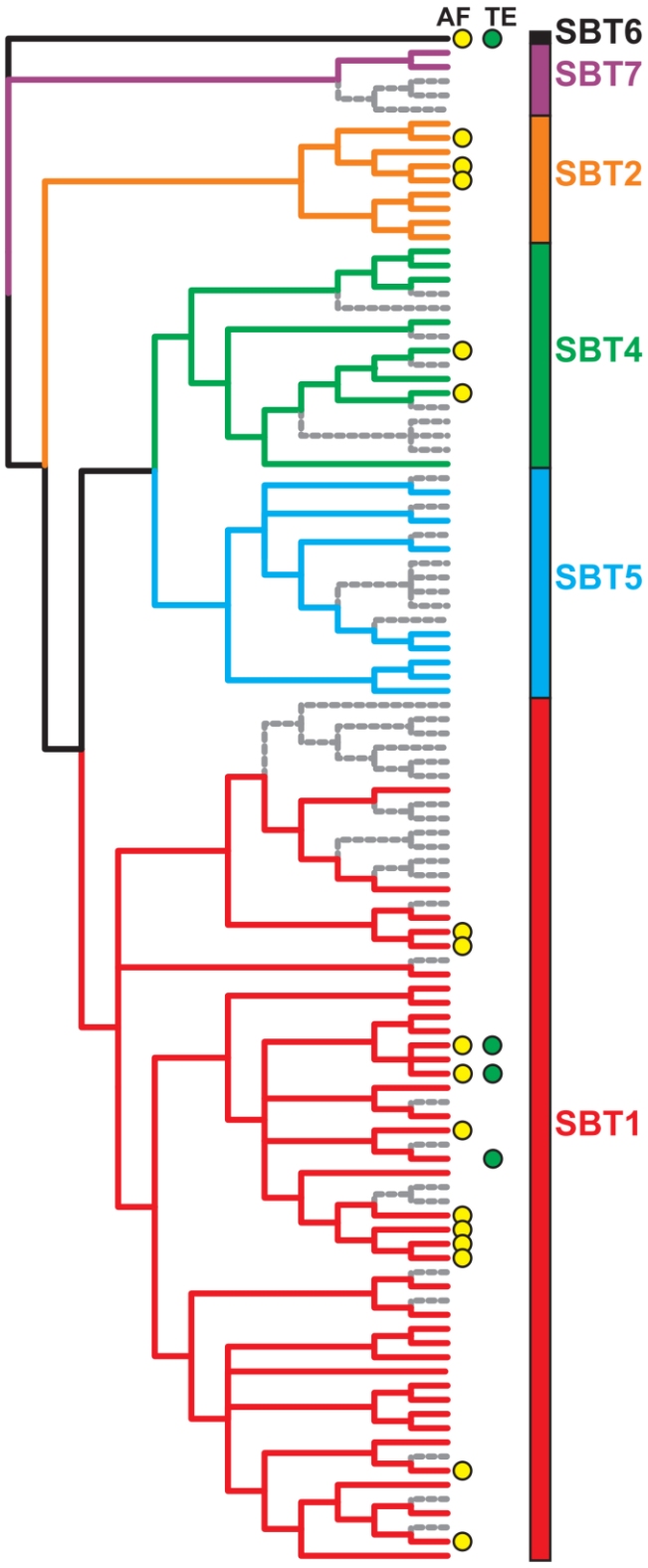


Figure 4. Birth and death of subtilase gene paralogs in *N. benthamiana.* The evolutionary history of the subtilisin gene family was inferred by using the Maximum Likelihood method based on the Whelan and Goldman model. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analysed. Putative pseudogenes are indicated in grey. Subtilases identified in apoplastic fluid (AF) and/or total extract (TE) are indicated with yellow and green dots, respectively. Naming of subtilase clades is according to (Taylor and Qiu, 2017). **Figure S2** includes the individual names.

During this analysis we discovered three subtilisin genes that are missing in our NbB database, and six incomplete sequences lacking 5-107 amino acids. In addition, five putative pseudo-genes were annotated as protein-encoding genes and were removed from the final NbC database, and 18 subtilase genes were found to be duplicated and these duplicates were removed in the final NbC database (**Table S3**). In comparison, the Niben1.0.1 genome annotation predicts 103 different subtilisin gene products. However, we found that these annotated genes correspond to 38 pseudo-genes and 49 protein-encoding genes - none of which are correctly annotated - while 16 subtilisin genes are absent from Niben1.0.1 (**Table S3**). Furthermore, the predicted proteome from the Nbv5.1 primary+alternate transcriptome contains more than 400 subtilisin gene products, largely due to a large number of chimeric sequences. In conclusion, the new database represents a significant improvement over previous genome annotations and facilitates more accurate and meaningful phylogenetic analysis of gene families in *N. benthamiana*.

Improved accuracy for genome editing: the subtilase gene-family

Target selection for genome editing is improved by the use of our new database for several reasons: 1) gene-models in this database are more complete; 2) fewer pseudo-genes are annotated as protein-encoding genes; 3) gene duplication is reduced as compared to the de-novo transcriptome; and 4) the remaining duplication in our database is easily resolved for genes of interest as it mostly involves genes with slight sequence variations between the different draft genomes. These sequence variations may be due to heterozygosity or technical artefacts of the sequencing and assembly. As an example we show the gene-model of one of the subtilisins in the different databases. In our NbC database, this subtilisin is encoded by a single-exon gene-model of 2,268bp encoding for a 756 amino acid protein (**Figure 5a**). This subtilisin is highly fragmented in the Niben1.0.1 genome assembly, with parts of the sequence

present on different contigs, while the gene is only partially annotated (**Figure 5b**). The last 90bp of this gene are not annotated in the Nbv0.5 genome (**Figure 5c**). Furthermore, there is a 132bp insertion in the Niben0.4.4 genome assembly resulting in a predicted protein with a 44 amino-acid insertion (**Figure 5d**). Additionally, we identified 13 sequences corresponding to partial or chimeric variants of this subtilisin are present in the Nb5.1 primary + alternate predicted proteome using BLAST with no full match. Finally, this subtilisin differs by three non-synonymous SNPs between the Niben1.0.1 and Nbv0.5 genome assemblies, while two of these non-synonymous SNPs are present in the Niben0.4.4 genome assembly (**Figure 5b,d**). In conclusion, this example displays how combining gene-models derived from different genome assemblies has made our database more complete than annotating any single genome assembly currently available.

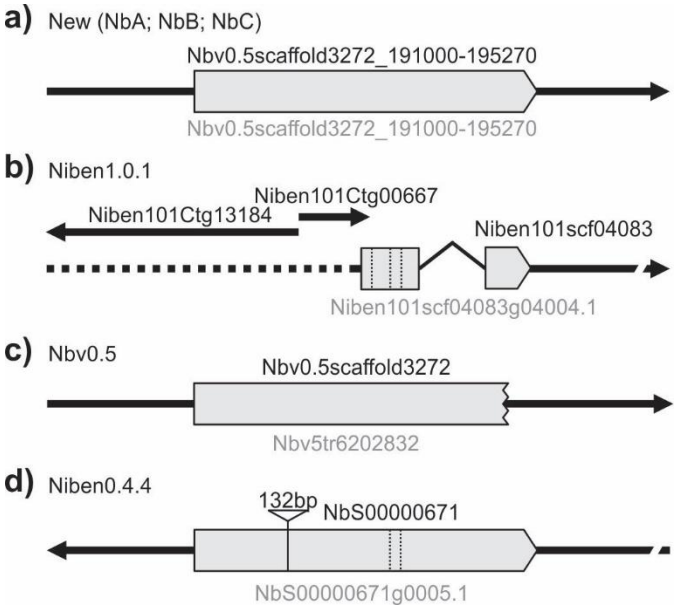


Figure 5. Example of a subtilase gene misannotation in different genome assemblies. The gene-models corresponding to a subtilase encoded by Nbv0.5scaffold3272_191000-195270 were retrieved from the different databases: **a)** NbC, **b)** Niben1.0.1 genome, **c)** Nbv0.5 genome, and **d)** Niben0.4.4 genome. CDS is annotated in grey boxes. The dotted line indicates a stretch of unknown sequence (Ns). Vertical dotted lines indicate SNPs, and the triangle indicate an insertion. Strand direction is indicated by arrows.

Although our NbC database does not contain the genomic context and lacks non-coding genes, this database will vastly improve research on *N. benthamiana*. We trust our NbC database to be useful for the large research community of plant scientists using *N. benthamiana* as a model system, for example to identify novel interactors in Co-IP experiments, but also to facilitate reverse genetic approaches such as genome editing and VIGS.

Material & Methods

Sequence retrieval - The predicted proteomes for *N. attenuata* (GCF_001879085.1) (Xu *et al.*, 2017) (http://nadh.ice.mpg.de/NaDH/), *N. tabacum* TN90 (GCF_000715135.1) (Sierro *et al.*, 2013), *N. sylvestris* (GCF_000393655.1) (Sierro *et al.*, 2013) and *N. tomentosiformis* (GCF_000390325.2) (Sierro *et al.*, 2013), and *Daucus carota* subsp. *sativus* (GCA_001625215.1) (Iorizzo *et al.*, 2016) were downloaded from Genbank. In addition, we retrieved 565 full-length *N. benthamiana* protein sequences from Genbank. The *Arabidopsis thaliana* predicted proteome (Araport11_genes.201606.pep) was obtained from Araport (Cheng *et al.*, 2017). The *Solanum melongena* predicted proteome (SME_r2.5.1_pep) was obtained from the Eggplant Genome DataBase (Hirakawa *et al.*, 2014). The *N. obtusifolia* (NIOBT_r1.0) predicted proteome was obtained from the *Nicotiana attenuata* Data Hub (Xu *et al.*, 2017) (http://nadh.ice.mpg.de/NaDH/). The *Petunia axillaris* N (Petunia_axillaris_v1.6.2_proteins) and *P. inflata* S6 (Petunia_inflata_v1.0.1_proteins) (Bombarely *et al.*, 2016), *Capsicum annuum glabriusculum* (CaChiltepin.pep) and *C. annuum zunla-1* (CaZL1.pep) (Qin *et al.*, 2014), *C. annuum* cv CM334 (Pepper.v.1.55.proteins.annotated) (Kim *et al.*, 2014), *Solanum tuberosum* (PGSC_DM_v3.4_pep) (Consortium, 2011), and *Solanum lycopersicum* (ITAG3.2_proteins) (Consortium, 2012) predicted proteomes were downloaded from Solgenomics. The *N. benthamiana* draft genome builds Niben1.0.1 and Niben0.4.4 - both generated by the Boyce

Thompson Institute for Plant Research (BTI) (Bombarely *et al.*, 2012) - were downloaded from Solgenomics, and the Nbv0.5 and Nbv0.3 draft genomes were made available by the Waterhouse lab at the Queensland University of Technology (Naim *et al.*, 2012).

Annotation - In order to extract gene-models from the published *N. benthamiana* draft genomes we combined all the *Nicotiana* protein sequences, except for those from *N. obtusifolia*, in one database, with the addition of 110 genes which we had previously manually curated leading to a database with 226,543 protein sequences. We used CD-HIT (v4.6.8) (Fu *et al.*, 2012) to cluster these sequences at a 95% identity threshold and reduce the redundancy in our database while removing partials (Nicotiana_db95; 85,453 sequences). This database was used to annotate the gene-models in the different *N. benthamiana* genome builds using Scipio version 1.4.1 (Keller *et al.*, 2008) which was run with default settings. After running Scipio we used Augustus (v3.3) (Stanke *et al.*, 2006) to extract complete and partial gene models. Putative pseudo-genes (containing internal stop codons) and genes lacking an ATG start or stop codon were stored separately. Transdecoder (v5.0.2) (Haas *et al.*, 2013) was used to retrieve the single-best ORF on the putative pseudo-genes containing homology to the Nicotiana_db95 database as determined by BlastP searches. If a putative pseudo-gene contained an ORF >90% of the annotated gene length and lacking <30 amino acids it was considered a putative gene. Other putative pseudo-genes were discarded. Next we used CD-HIT-EST to filter the redundancy from this database. First, we used CD-HIT-EST to cluster the CDS derived from the gene-models derived from the different databases at 100% identity. Next, we selected the longest sequence at 99% identity between the different genome builds using CD-HIT-EST-2D in the following order for both the complete and the partial databases: Niben1.0.1 > Nbv0.5 > Niben0.4.4 > Nbv0.3. Since sequences which are smaller are maintained like this we used the reduced databases in the opposite direction to remove partial genes: Nbv0.3 > Niben0.4.4 > Nbv0.5 > Niben1.0.1. Finally we used CD-HIT-EST-2D to

remove genes from the partial database with a longer representative in the complete database at 99% identity and vice-versa. This resulted in the NbA database. We compared this database for proteomic analysis on the described proteomics dataset containing apoplastic fluid (AF) samples and full-leaf extract (TE) samples and compared its performance to the other published predicted proteomes. For this analysis we predicted the Nbv5.1 proteome from the transcriptome using Transdecoder and selecting the single-best ORF with homology to the Nicotiana_db95 database, and filtered the database using CD-HIT at 100% identity. Proteins for which peptides were identified in the other databases but absent from the NbA database search were extracted, clustered at 100% using CD-HIT, and re-annotated in the genomes as above. This resulted in the NbB database. Finally, we ran BUSCO (v3.0.2; dependencies: NCBI-BLAST v2.7.1+; HMMER v3.1; Augustus v3.3) (Simão *et al.*, 2015; Waterhouse *et al.*, 2018) on the different *N. benthamiana* predicted proteomes using the plants set (Embryophyta_odb9), extracted the missing BUSCOs and re-annotated these as above. Additionally, we manually inspected the database for the PLCP, subtilisin, VPE, and GH35-domain encoding gene families, and manually removed redundant sequences. This resulted in our final database. This database was annotated using SignalP (v4) (Petersen *et al.*, 2011), ApoplastP (v1.0.1) (Sperschneider *et al.*, 2018), and PFAM (v31) (Finn *et al.*, 2016). Finally we annotated the predicted proteome with GO terms and UniProt identifiers using Sma3s v2 (Casimiro-Soriguer *et al.*, 2017).

Sample preparation for proteomics and definition of biological replicates - Four-week old *N. benthamiana* plants were used. The AF was extracted by vacuum infiltrating *N. benthamiana* leaves with ice-cold MilliQ. Leaves were dried to remove excess liquid, and apoplastic fluid was extracted by centrifugation of the leaves in a 20 ml syringe barrel (without needle or plunger) in a 50 ml falcon tube at 2000 x g, 4°C for 25 min. Samples were snap-frozen in liquid nitrogen and stored at -80°C prior to use. TE was collected by removing the

central vein and snap-freezing the leaves in liquid nitrogen followed by grinding in a pestle and mortar and addition of three volumes of phosphate-buffered saline (PBS) (w/v). One biological replicate was defined as a sample, AF or TE, consisting of one leaf from three independent plants (3 leaves total). Four independent biological replicates were taken for AF and TE.

Protein digestion and sample clean-up - AF and TE sample corresponding to 15 µg of protein was taken for each sample (based on Bradford assay). Dithiothreitol (DTT) was added to a concentration of 40 mM, and the volume adjusted to 250 µl with MS-grade water (Sigma). Proteins were precipitated by the addition of 4 volumes of ice-cold acetone, followed by a 1 hr incubation at -20°C and subsequent cetrifugation at 18,000 x g, 4°C for 20 min. The pellet was dried at room temperature (RT) for 5 min and resuspended in 25 µL 8 M urea, followed by a second chloroform/methanol precipitation. The pellet was dried at RT for 5 min and resuspended in 25 µL 8 M urea. Protein reduction and alkylation was achieved by sequential incubation with DTT (final 5 mM, 30 min, RT) and iodoacetamide (IAM; final 20 mM, 30 min, RT, dark). Non-reacted IAM was quenched by raising the DTT concentration to 25 mM. Protein digestion was started by addition of 1000 ng LysC (Wako Chemicals GmbH) and incubation for 3 hr at 37°C while gently shaking (800 rpm). The samples were then diluted with ammoniumbicarbonate (final concentration 80 mM) to a final urea concentration of 1 M. 1000 ng Sequencing grade Trypsin (Promega) was added and the samples were incubated overnight at 37°C while gently shaking (800 rpm). Protein digestion was stopped by addition of formic acid (FA, final 5% v/v). Tryptic digests were desalted on home-made C18 StageTips (Rappsilber *et al.*, 2007) by passing the solution over 2 disc StageTips in 150 µL aliquots by centrifugation (600-1200 x g). Bound peptides were washed with 0.1% FA and subsequently eluted with 80% Acetonitrile (ACN). Using a vacuum concentrator (Eppendorf) samples were dried, and the peptides were resuspended in 20 µL 0.1% FA solution.

LC-MS/MS - The samples were analysed as in (Grosse-Holz *et al.*, 2018). Briefly, samples were run on an Orbitrap Elite instrument (Thermo) (Michalski *et al.*, 2011) coupled to an EASY-nLC 1000 liquid chromatography (LC) system (Thermo) operated in the one-column mode. Peptides were directly loaded on a fused silica capillary (75 µm × 30 cm) with an integrated PicoFrit emitter (New Objective) analytical column packed in-house with Reprosil-Pur 120 C18-AQ 1.9 µm resin (Dr. Maisch), taking care to not exceed the set pressure limit of 980 bar (usually around 0.5-0.8 µl/min). The analytical column was encased by a column oven (Sonation; 45°C during data acquisition) and attached to a nanospray flex ion source (Thermo). Peptides were separated on the analytical column by running a 140 min gradient of solvent A (0.1% FA in water; ; Ultra-Performance Liquid Chromatography (UPLC) grade) and solvent B (0.1% FA in ACN; UPLC grade) at a flow rate of 300 nl/min (gradient: start with 7% B; gradient 7% to 35% B for 120 min; gradient 35% to 100% B for 10 min and 100% B for 10 min) at a flow rate of 300 nl/min.). The mass spectrometer was operated using Xcalibur software (version 2.2 SP1.48) in positive ion mode. Precursor ion scanning was performed in the Orbitrap analyzer (FTMS; Fourier Transform Mass Spectrometry) in the scan range of m/z 300-1800 and at a resolution of 60000 with the internal lock mass option turned on (lock mass was 445.120025 m/z, polysiloxane) (Olsen *et al.*, 2005). Product ion spectra were recorded in a data-dependent manner in the ion trap (ITMS) in a variable scan range and at a rapid scan rate. The ionization potential was set to 1.8 kV. Peptides were analysed by a repeating cycle of a full precursor ion scan (1.0 x 106 ions or 50ms) followed by 15 product ion scans (1.0×10^4 ions or 50ms). Peptides exceeding a threshold of 500 counts were selected for tandem mass (MS2) spectrum generation. Collision induced dissociation (CID) energy was set to 35% for the generation of MS2 spectra. Dynamic ion exclusion was set to 60 seconds with a maximum list of excluded ions consisting of 500 members and a repeat count of one. Ion injection time prediction, preview mode for the Fourier transform mass spectrometer (FTMS, the orbitrap),

monoisotopic precursor selection and charge state screening were enabled. Only charge states higher than 1 were considered for fragmentation.

Peptide and Protein Identification - Peptide spectra were searched in MaxQuant (version 1.5.3.30) using the Andromeda search engine (Cox *et al.*, 2011) with default settings and label-free quantification and match-between-runs activated (Cox and Mann, 2008; Cox *et al.*, 2014) against the databases specified in the text including a known contaminants database. Included modifications were carbamidomethylation (static) and oxidation and N-terminal acetylation (dynamic). Precursor mass tolerance was set to ±20 ppm (first search) and ±4.5 ppm (main search), while the MS/MS match tolerance was set to ±0.5 Da. The peptide spectrum match FDR and the protein FDR were set to 0.01 (based on a target-decoy approach) and the minimum peptide length was set to 7 amino acids. Protein quantification was performed in MaxQuant (Tyanova *et al.*, 2016), based on unique and razor peptides including all modifications.

Proteomics processing in R - Identified protein groups were filtered for reverse and contaminants proteins and those only identified by matching, and only those protein groups identified in 3 out of 4 biological replicates of either AF or TE were selected. The LFQ values were $\log_2$ transformed, and missing values were imputed using a minimal distribution as implemented in imputeLCMD (v2.0) (Lazar, 2015). A moderated t-test was used as implemented in Limma (v3.34.3) (Ritchie *et al.*, 2015; Phipson *et al.*, 2016) and adjusted using Benjamini–Hochberg (BH) adjustment to identify protein groups significantly differing between AF and TE. Bonafide apoplastic protein groups were those only detected in AF and those significantly ($p \leq 0.01$) $\log_2$ fold change ≥ 1.5 in AF samples. Protein groups only detected in TE and those significantly ($p \leq 0.01$) $\log_2$ fold change ≤ -1.5 depleted in AF samples were considered intracellular. The remainder was considered both apoplastic and intra-cellular. Majority proteins were annotated with SignalP, PFAM, MEROPS (v12) (Rawlings *et al.*,

2018), GO, and UniProt keywords identifiers. A BH-adjusted Hypergeometric test was used to identify those terms that were either depleted or enriched ($p \leq 0.05$) in the bonafide AF protein groups as compared to bonafide AF depleted proteins or protein groups present both in the AF and TE.

Phylogenetic analysis - Predicted proteomes were annotated with PFAM identifiers, and all sequences containing a Peptidase S8 (PF00082) domain were extracted from the different databases. Additionally, we manually curated the subtilisin gene-family in the Niben1.0.1 draft genome, identifying putative pseudo-genes which were annotated as protein-encoding genes, as well as missing genes and incorrect gene models or genes in which the reference sequence was absent in Niben1.0.1. Tomato subtilisins were retrieved from Solgenomics, and other previously characterized subtilisins (Taylor and Qiu, 2017) were retrieved from NCBI. Clustal Omega (Sievers *et al.*, 2011; Li *et al.*, 2015) was used to align these sequences. The putative pseudo-gene sequences were substituted with the best blast hit in NCBI in order to visualize pseudogenization in the alignment and phylogenetic tree. Determining the best model for maximum likelihood phylogenetic analysis and the phylogenetic analysis was performed in MEGA X (Kumar *et al.*, 2018). The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan and Goldman model. A discrete Gamma distribution was used to model evolutionary rate differences among sites, and the rate variation model allowed for some sites to be evolutionarily invariable. All positions with less than 80% site coverage were eliminated. Niben101Scf00595_742942-795541 was used to root the phylogenetic trees.

Data Availability - The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository (https://www.ebi.ac.uk/pride/archive/) with the data set identifier PXD010435. During the review process the data can be accessed via a reviewer account (Username:

reviewer17475@ebi.ac.uk; Password: PQSfFZyN). Samples FGH01-04 represent AF and FGH05-08 represent TE.

Acknowledgements. We would like to thank Philippe Varennes-Jutras and Daniela Sueldo for critically reading the manuscript and providing important suggestions for improving the manuscript.

Funding. This work has been supported by 'The Clarendon Fund' (JK, FH), and the ERC Consolidator grant 616449 'GreenProteases' (RvdH, FGH, JK). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests. The authors have declared that no competing interests exist.

Author contributions. Conceptualization: JK, RvdH; Formal analysis: JK; Funding acquisition: RvdH; Wetlab experiments: FHG; Proteomics: FK, MK; Programming: JK, FH; Writing: JK, RvdH.

Supplemental figures

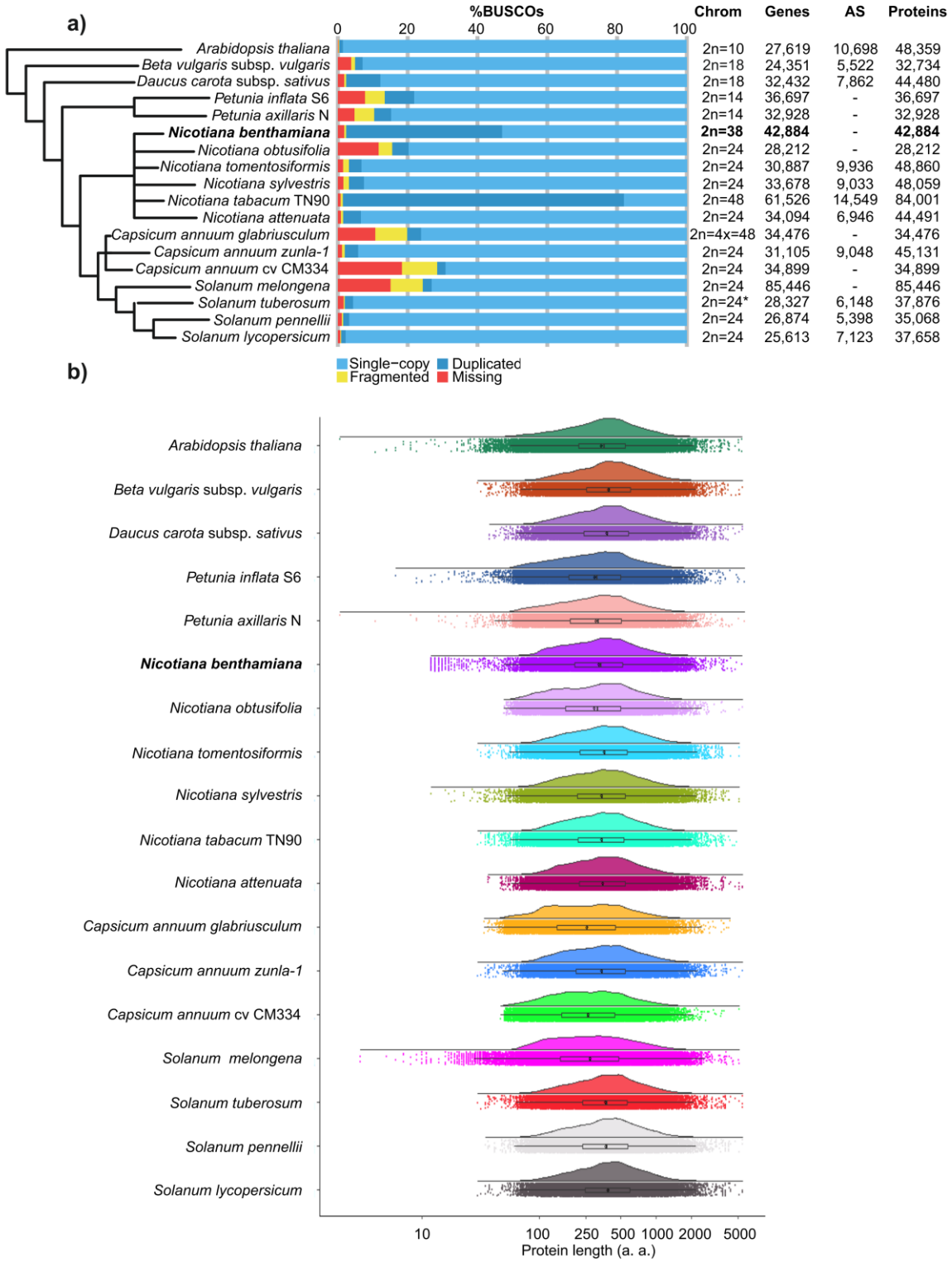


Figure S1: Comparison of Solanaceae proteomes. **a)** Completeness of the predicted proteomes from sequenced Solanaceae genomes was estimated using BUSCO v3 with the embryophyta database. *Certain *Solanum tuberosum* species are polyploid. **b)** Violin and boxplot g of $\log_{10}$ protein length distribution of each predicted proteome. Jittered dots show the raw underlying data.

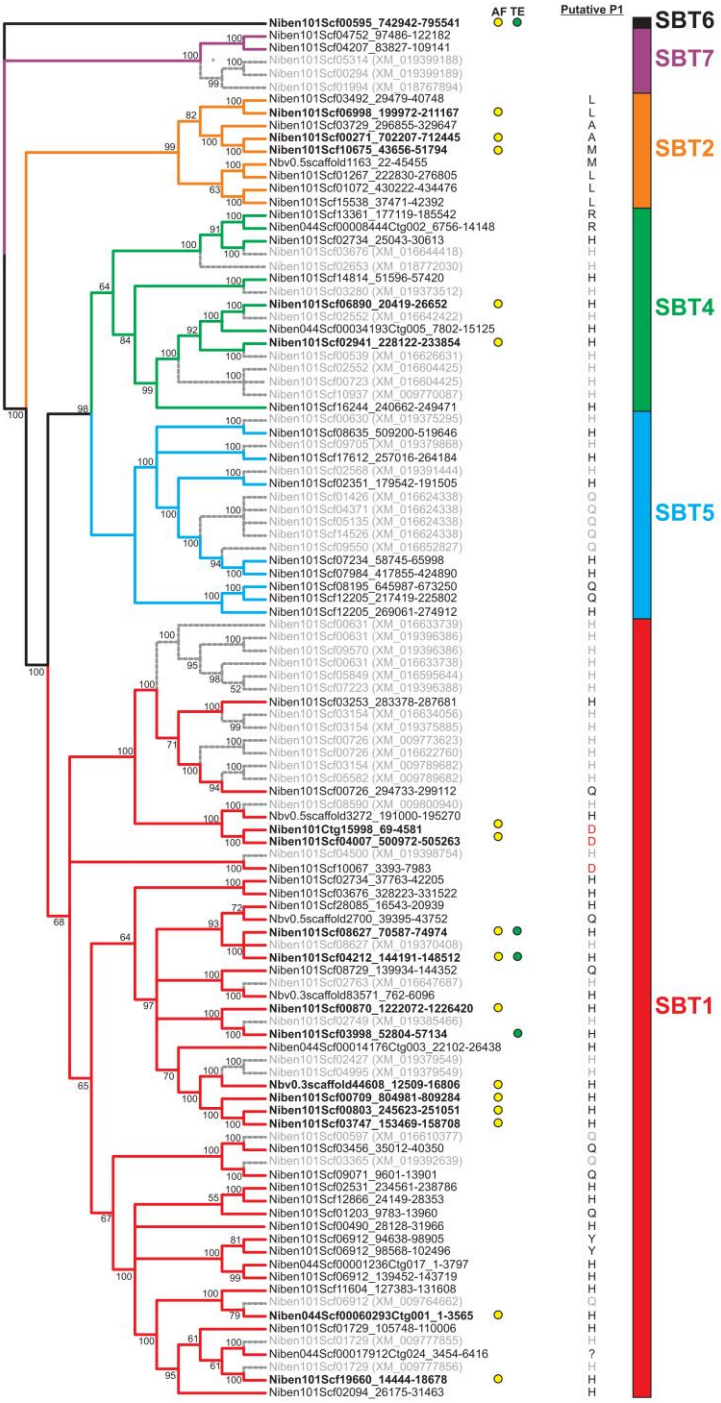


Figure S2: Phylogenetic analysis of the subtilisin gene-family from figure 4 with names. The evolutionary history of the subtilase gene family was inferred by using the Maximum Likelihood method based on the Whelan and Goldman model. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analysed. Putative pseudogenes are indicated in grey. Subtilases identified in apoplastic fluid (AF) and/or total extract (TE) are indicated with yellow and green dots, respectively. Putative P1 based on residue at the prodomain junction is indicated, putative phytaspases are indicated in red. Naming of subtilase clades according to (Taylor and Qiu, 2017).

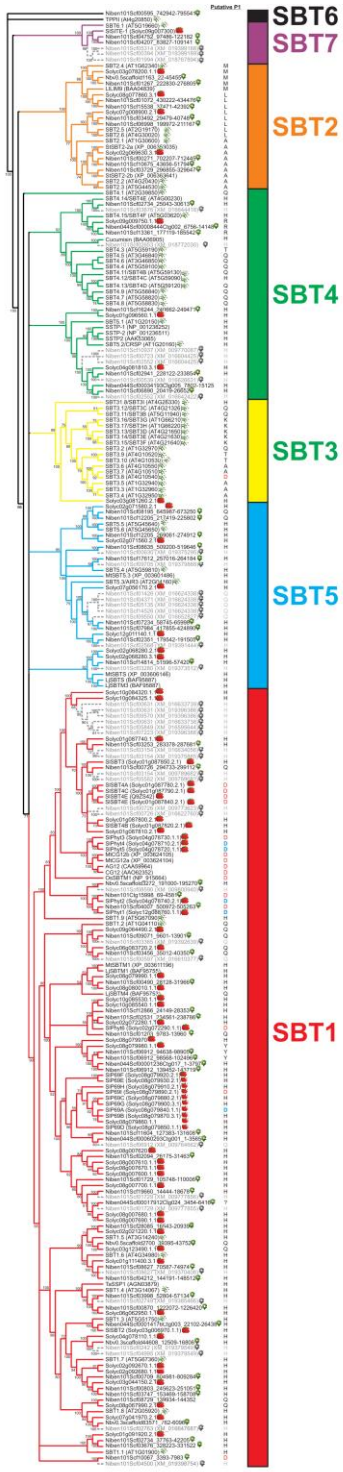


Figure S3: Phylogenetic analysis of the subtilisin gene-family of tomato and Arabidopsis and including other previously characterized subtilisins. The evolutionary history of the subtilase gene family was inferred by using the Maximum Likelihood method based on the Whelan and Goldman model. The bootstrap consensus tree inferred from 250 replicates is taken to represent the evolutionary history of the taxa analysed. Putative pseudogenes are indicated in grey. Putative P1 based on residue at the prodomain junction is indicated. Putative phytaspases are indicated in red, confirmed phytaspases in blue. Naming of subtilase clades according to (Taylor and Qiu, 2017).

Other supplemental datasets:

Table S1: GO-SLIM term enrichment complete at $p \leq 0.05$

Table S2: MEROPS family term enrichment complete

Table S3: Gene-model comparison

Supplemental dataset 1: NbC gene-models database fasta nucleotide sequence

Supplemental dataset 2: NbC gene-models database gff3 annotation

Supplemental dataset 3: NbC predicted CDS

Supplemental dataset 4: NbC predicted proteome

Supplemental dataset 5: PFAM, SignalP, ApoplastP, and Sma3 v2 annotation of the NbC predicted proteome

Chapter 7

General Discussion and conclusion

The key messages from this thesis

The work presented in this thesis contributes to the understanding of the molecular mechanisms involved in pathogen recognition, and the evolution of these mechanisms. I believe that evolutionary analyses of resistance (*R*) genes may not only shed light on how these genes came about and how they function, but also provide crucial information for rationally engineering them. Indeed, "*Nothing in biology makes sense except in the light of evolution*" (Dobzhansky, 1973).

In the remainder of this discussion I will comment on the three key messages from my thesis: i) the description of the molecular mechanisms by which *R* genes function, ii) the evolution of guarded decoys and their receptors, and iii) decoy swapping as an approach to engineering of altered recognition spectra. Finally, I will discuss the novel *Nicotiana benthamiana* gene-model annotation, and how these gene-models can enhance plant-science using this model organism.

The molecular mechanisms of R genes

In **chapter 2** I describe the molecular mechanisms by which *R* genes function. *R* genes in the broadest definition are genes with a role in innate immunity. Several models have been proposed to explain the evolution and mechanisms involved in the plant innate immune system. Arguably the most prominent model is the zigzag model proposed by Jones and Dangl (Jones and Dangl, 2006). Here the authors distinguish between pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) and effector-triggered immunity (ETI). While this paper has served as an important framework for thinking about plant-pathogen interactions, it does not distinguish between evolutionary processes or molecular mechanisms. Additionally, there is no experimental evidence showing that PTI and ETI are distinct molecular processes, evolving in distinct manners. Other models have sought to redefine the plant innate immune system, avoiding the terms PTI and ETI. Cui *et al.* for example propose to view the plant

immune system as a surveillance system to detect invasion (Cui *et al.*, 2015). However, this view does not consider loss-of-susceptibility mechanisms of resistance.

We took a different approach, in which we conducted a comprehensive meta-analysis of the molecular mechanisms which underlie *R* gene function, thereby uncovering nine general mechanisms. It is important to note, however, that the exact details are likely to be different for each *R* gene. Additionally, not every *R* gene will be easy to classify according to these proposed mechanisms, especially *R* genes which function by loss-of-susceptibility mechanisms.

An example of a *R* gene which is difficult to place is tomato (*Solanum lycopersicum* L.) *Asc-1*. *Asc-1* prevents the disruption of sphingolipid metabolism by the necrotrophic fungal pathogen *Alternaria alternata* f. sp. *lycopersici* (Spassieva *et al.*, 2002). *A. alternata* produces the host-selective toxin AAL, which kills the host cell by inducing programmed cell-death. *Asc-1* encodes for an enzyme involved in sphingolipid metabolism, and upon treatment with AAL, *Asc1*-lines produce different sphingolipids, while *asc-1*-lines respond by producing less sphingolipids (Spassieva *et al.*, 2002). *Asc-1* could act either by i) a loss-of-interaction manner, not being able to interact with AAL, ii) an active manner in which the *Asc-1* allele is more active than the *asc-1* allele, or iii) a host reprogramming manner in which AAL induces the production of different sphingolipids by *Asc-1*. These mechanisms cannot be separated without further studies into the mechanism by which AAL toxin works. This knowledge is not relevant for the use of *Asc-1* in the field. However, for engineering resistance to *A. alternata* in different host species the molecular mechanism is relevant. By studying the evolution of *Asc-1*, combined with a better structural understanding of this interaction it could be possible to engineer this trait into other species. Similarly difficult to classify is $Hs1^{pro-1}$, which provides nematode (*Heterodera schachtii* Schmidt) resistance in sugarbeet (*Beta vulgaris* L.) (Cai *et al.*, 1997). Originally it was thought that $Hs1^{pro-1}$ encodes for a leucine-rich-repeat (LRR) containing receptor-like protein (RLP) (Cai *et al.*, 1997). Homology modelling of $Hs1^{pro-1}$,

however, reveals a potential structural homology to putative 2,3-dioxygenase-like enzymes acting on indolic compounds. Additionally, *Hs1^{pro-1}*-like genes are conserved in plants. The resistance causing *Hs1^{pro-1}* gene-product appears to miss the N-terminal Hs1^{pro-1} domain, and only consists of the C-terminal Hs1^{pro-1} domain. Whether this would result in a lack of activity or an altered activity is unknown. Since *Hs1^{pro-1}* is a conserved gene, and these nematodes are general pathogens, it is possible that this mechanism could be transferred to other species. A third *R* gene with an unknown molecular mechanism is the barley (*Hordeum vulgare*) *Rpg1* gene which confers resistance to the stem rust pathogen *Puccinia graminis* f. sp. *tritici*. *Rpg1* encodes for a receptor-like kinase which is specifically phosphorylated and subsequently degraded by the concerted action of two fungal effectors which can interact with Rpg1 (Nirmala *et al.*, 2011). This ultimately results in a hypersensitive response (HR) and resistance to *P. graminis*. Whether this is an active response or whether Rpg1 is guarded by another receptor, like Arabidopsis (*Arabidopsis thaliana*) RIN4 or tomato Pto, is unknown. Alternatively, Rpg1 may belong to an entirely novel class of intracellular receptors involved in pathogen perception. Indeed, recently it was shown that *Yr15*, a *R* gene originating from wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*) which provides broad-resistance to yellow rust caused by *Puccinia striiformis* f. sp. *tritici*, encodes for a protein with a similar domain architecture (Klymiuk *et al.*, 2018). This strongly hints at the possibility that this represents a novel class of immune receptors.

We also did not include the recognition of danger-associated molecular patterns (DAMPs) as a mechanism. DAMPs are endogenous signals which are produced or released upon cellular stress, cellular damage, or non-physiological cell-death. DAMPs can be involved in either triggering or reinforcing immune signalling and may thereby facilitate indirect perception of pathogens. Furthermore, DAMPs can function to reinforce immune signalling, or as short-distance signals to prime neighbouring cells for an incoming pathogen.

For example, Arabidopsis can perceive pectin derivatives, thereby triggering immune signalling. These derivatives are produced when pathogen polygalacturonases (PGs) are partially inhibited by plant-derived PG Inhibitor Proteins (PGIPs) (Lorenzo *et al.*, 1994). This partial inhibition leads to the generation of immunogenic oligogalacturonides (OGs) from pectin (Benedetti *et al.*, 2015), which are perceived by the Arabidopsis Wall-Associated receptor Kinase 1 (WAK1), triggering an immune response (Brutus *et al.*, 2010). Other DAMPs are involved in fortifying the immune response. Arabidopsis prePIP1 is transcriptionally upregulated upon pathogen perception, and the secreted PIP1 peptide is perceived by RLK7, thereby amplifying flagellin-triggered immune signalling (Hou *et al.*, 2014). Arabidopsis PROPEP peptides are another group of peptide DAMPs, collectively known as PEPs, which lack a signal peptide (Huffaker *et al.*, 2006). PEPs are perceived by the PEPR1/PEPR2 RLKs (Tang *et al.*, 2015), which upon PEP recognition trigger immune responses with an output similar to that triggered by flagellin. Interestingly, BAK1-depletion sensitizes PEPR-signalling for a pro-death role, leading to increased basal resistance against virulent (hemi-)biotrophic pathogens (Yamada *et al.*, 2016).

These three examples point to a role for DAMPs in the initial perception of pathogens (OGs), and in amplifying the initial immune response (PIP1/PEPs). Furthermore, DAMPs could maintain the immune response, once effectors start targeting (co-)receptors involved in the initial perception such as BAK1 (PEPs). However, to be classified as *R* genes it must be shown that the DAMP-encoding genes and their receptors underlie polymorphic (quantitative) resistance to pathogens, or that their transfer between species can enhance resistance in a DAMP-specific manner. This might be the case for *DORN1*. Arabidopsis *DORN1* encodes for a lectin domain containing receptor-like kinase (RLK) which recognizes extracellular ATP (Choi *et al.*, 2014). DORN1 is a virulence target of the *Phytophthora infestans* effector IPI-O (Bouwmeester *et al.*, 2011). Arabidopsis *dorn1* knock-out lines are more susceptible to *P.*

infestans (Bouwmeester *et al.*, 2011) and *Pseudomonas syringae* pv. *tomato* DC3000 (Balagué *et al.*, 2017). Overexpression of Arabidopsis *DORN1* in potato (*Solanum tuberosum*) and *Nicotiana benthamiana* results in enhanced resistance to *P. infestans* (Bouwmeester *et al.*, 2014). However, whether this is specific to enhanced DAMP signalling is unknown.

This points to another difficulty: enhanced resistance or enhanced susceptibility upon overexpression or knock-out of putative *R* genes is not sufficient to prove a causal role in disease resistance under natural conditions, because the observed effects may be indirect or unspecific. This concept is very elegantly demonstrated in a recent paper in which the network of Arabidopsis LRR-RLKs is described (Smakowska-Luzan *et al.*, 2018). By removing an important component of a sub-network, the authors show that the overall network stability is altered, and thereby the activity of other indirectly connected parts of the network are altered (Smakowska-Luzan *et al.*, 2018). Knock-out or overexpression of putative *R* genes may similarly alter network stability, leading to incorrect conclusions about their function. These networks may in fact be so interconnected that any gene which is expressed in a disease-relevant context may affect the function of core disease-related genes (Boyle *et al.*, 2017). Going forward it is clear that the biochemical understanding of gene networks is crucial for the interpretation of omics-type data. Additionally, understanding how these biochemical networks evolve may shed light on the molecular mechanisms by which they function.

In conclusion, the nine described molecular mechanisms provide a framework to think about molecular mechanisms by which *R* genes could function. Much research has focused on identifying novel *R* genes. The advance of next-generation sequencing approaches is likely to speed up the rate of *R* gene discovery. However, if we want to rationally engineer novel resistance traits, it is important not just to discover more genes, but also to understand how these genes function. This paper provides a step in that direction summarizing the known mechanisms. I expect many more mechanisms to exist in plants. Indeed, during the publication

of this chapter, a pre-print became available describing an extension of the NLR-ID mechanism (Fujisaki *et al.*, 2017). Here the authors describe the interaction of an NLR through its integrated NOI domain with another host protein, OsExo70-F3, which in turn interacts with the *Magnaporthe oryzae* effector AvrPii (Fujisaki *et al.*, 2017). This is the first example of an NLR-ID indirectly recognizing a pathogen effector through its integrated domain.

Understanding the evolution of R genes may shed light on their mechanisms

In **chapter 3** I describe the evolution *Cf-2* and *Rcr3*. Previously it was hypothesized that Rcr3 acts as a decoy for the recognition of Avr2 by *Cf-2*, and that Pip1 is the operative target of Avr2 (van der Hoorn and Kamoun, 2008). According to this hypothesis *Rcr3* would have specifically evolved for this purpose. However, this seems highly unlikely since both *Rcr3* and *Pip1* are ancient genes, conserved in Solanaceae, whereas *Cf-2* is a tomato-specific gene. Additionally, it seems unlikely that *Rcr3* acts as a general co-receptor for this gene family, since *Cf-5* does not require *Rcr3* for the recognition of Avr5, nor does it require *Pip1* or *Pfp1*.

What the role of Rcr3 in immunity is in the absence of *Cf-2* is currently unknown. Both Rcr3 and Pip1 activity is increased upon signalling associated with biotic stress (Shabab *et al.*, 2008). *Pip1* has a basal role in immunity against a variety of pathogens, whereas *Rcr3* is involved in basal resistance against *P. infestans* (Ilyas *et al.*, 2015; Song *et al.*, 2009). It may very well be that Rcr3 is not an operative virulence target of Avr2 in the tomato-*Cladosporium fulvum* interaction. Alternatively, it may be that Rcr3 has not been identified as an operative virulence target because it functions in specific conditions, in a redundant manner, or in a minor way. From the perspective of molecular recognition in concert with *Cf-2*, it is not relevant whether Rcr3 is an operative virulence target or not.

Cf-2 was originally introgressed from the wild currant tomato (*Solanum pimpinellifolium*). It is possible that *Cf-2* is more wide-spread in tomato species. *Cf-4* and *Cf*-9, for example, are

maintained by balancing selection in different tomato species (Kruijt *et al.*, 2005). This indicates that *Cf-4* and *Cf-9* predate the speciation of tomato species, and that *C. fulvum* is an ancient pathogen of tomato (Kruijt *et al.*, 2005). Alternatively, this could indicate gene-flow between these species in the wild, or that these effectors may be conserved between different pathogens. Indeed, while many effectors are pathogen-, or even race/pathovar-specific, Avr4 appears to be a "core"-effector, conserved in Dothideomycete fungi (Stergiopoulos *et al.*, 2010). The Dothideomycetes class of fungi contains many plant-pathogenic species with diverse lifestyles and a broad range of monocot and dicot hosts (de Wit *et al.*, 2012; Ohm *et al.*, 2012). Avr4 homologs from diverse Dothideomycetes fungi can be recognized by *Cf-4* (Stergiopoulos *et al.*, 2010; de Wit *et al.*, 2012; Mesarich *et al.*, 2015). Whether *Cf-2* is similarly maintained is yet unknown. Both the *Hcr2* and *Hcr9* gene families are conserved in Solanaceae species. Since *C. fulvum* is only a pathogen of tomato species, it may be that homologs of these genes are involved in disease resistance against a wide variety of pathogens.

Finally, *Cf-2* exemplifies how *R* genes involved in indirect recognition of pathogens may evolve *de novo*, co-opting pre-existing components. Many other *R* gene products are involved in the indirect recognition of pathogens. However, how these genes have evolved has not been studied in-depth. It might be that novel *R* genes are more likely to indirectly recognize pathogens. Additionally, it may be that many *R* genes converge on guarding the same host components. For example, the barley Nod-like receptor (NLR) Pbr1 has independently converged on guarding a PBS1-like protein for the recognition of the *P. syringae* type III effector AvrPphB, similar to how the Arabidopsis NLR RPS5 guards PBS1 for the recognition of AvrPphB (Carter *et al.*, 2018). The same seems to be the case for RIN4, which is guarded by different NLRs in different species for the recognition of various effectors (Ashfield *et al.*, 2004; Kessens *et al.*, 2014; Selote and Kachroo, 2010). Therefore, it seems likely that common virulence targets are also often co-opted for the recognition of pathogens by independently

evolving *R* genes. It is tempting to speculate that Rcr3 and other apoplastic proteases may have independently been co-opted by additional *R* genes for the recognition of diverse pathogens.

Decoy swapping as a strategy to engineer novel recognition

In **chapter 5** I describe our efforts to engineer Pip1 for *Cf-2*-dependent recognition of Avr2. Pip1 can be engineered for the interaction with *Cf-2* with minimal changes. This engineering approach sheds light on the molecular mechanism by which *Cf-2* recognizes the interaction of Avr2 with Rcr3. We propose that Rcr3 and Cf-2 interact directly. Binding of Avr2 to Rcr3 is sufficiently strong to alter the interaction of Rcr3 and Cf-2, which allows the recruitment of BAK1, triggering downstream immune signalling. It is likely that Cf-2 is not specific for the recognition of Avr2. If indeed *Cf-2* is not specific to Avr2, it may be that engineering alternative decoys to interact with Cf-2 could result in an altered recognition spectrum. This is a hypothesis which will be addressed in future experiments. This decoy swapping approach may present a general approach for the engineering of altered pathogen recognition spectra.

The N. benthamiana apoplastic proteome

Finally, in **chapter 6** I describe the re-annotation of *N. benthamiana* gene-models, and the use of the predicted proteome derived from this dataset for the analysis of the apoplastic proteome. As expected, many different proteases are enriched in the apoplastic proteome. Also enriched are glycoside hydrolases. These enzymes are potentially involved in cell-wall remodelling and plant-pathogen interactions.

The original *N. benthamiana* genome annotation contains many mis-annotated gene-models. Accurately annotating gene-models in *N. benthamiana* is complicated due to the ancient allotetraploid origin of *N. benthamiana*, and the subsequent genomic rearrangements. This has resulted in the pseudogenization of many genes, including *N. benthamiana Rcr3* and *Pip1* (chapter 3). In the case of *NbPip1*, this pseudogenization appears to be very recent and specific

to the *N. benthamiana* isolate used as a model-organism. Therefore, it is likely that this pseudogenization is an ongoing process which has resulted in *N. benthamiana* being intermediate between a tetraploid and a diploid species regarding gene content. Comparative genomics of *N. benthamiana* with related species could provide interesting insights into the processes driving pseudogenization, genomic rearrangements, and genome reduction. Additionally, the use of different *N. benthamiana* ecotypes could be very useful for mapping traits of interest. A more contiguous *N. benthamiana* genome build is currently being pursued by several labs. The gene-models we describe will be useful for the annotation of these genome builds.

Conclusions

This thesis advances our knowledge on the molecular mechanisms of pathogen resistance. I highlight how one such mechanism has evolved, and how this knowledge can be used to engineer this mechanism. Novel *R* genes which encode for receptors evolve at the edges of signalling networks, but may co-opt highly interconnected protein hubs for the recognition of pathogens. A fundamental biochemical understanding of these networks is expected to give insights into the mechanisms of molecular recognition, as well as open new avenues for rational engineering of novel traits.

Appendix

Bodyguards: Pathogen-Derived Decoys That Protect Virulence Factors

Judith K. Paulus[1], Jiorgos Kourelis[1] and Renier A.L. van der Hoorn[1*]

[1]The Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, Oxford, UK

*Correspondence: renier.vanderhoorn@plants.ox.ac.uk (R.A.L. van der Hoorn).

Abstract

Recent studies on plant-pathogen interactions have exposed a new strategy used by plant pathogens: decoy effectors that protect virulence factors. Examples of these "bodyguards" include the recently discovered *Ps*XLP1 from *Phytophthora sojae* and truncated TALEs from *Xanthomonas oryzae*. These examples suggest important roles for seemingly non-functional effector proteins in distracting the host.

Main text

The Xyloglucan-specific Endo-Glucanase (*Ps*XEG1) acts as an important virulence factor of the soil-borne oomycete pathogen *Phytophthora sojae* (Ma *et al.*, 2015). During infection, *Ps*XEG1 is secreted into the apoplast of its host soybean (*Glycine max*) to macerate the host cell wall. However, *Ps*XEG1 is recognized by the plant's recognition machinery as a Pathogen-associated Molecular Pattern (PAMP), independent of its enzymatic activity (Ma *et al.*, 2015). *P. sojae* is able to suppress PAMP-triggered immune responses caused by *Ps*XEG1, presumably by injecting multiple intracellular RxLR effectors (Ma *et al.*, 2015). But in addition to being recognized as a PAMP, *Ps*XEG1 is targeted by a secreted endoglucanase inhibitor (*Gm*GIP1) of soybean. *Gm*GIP1 binds to *Ps*XEG1 and strongly inhibits XEG1-catalyzed depolymerisation of xyloglucan, reducing the virulence role of *Ps*XEG1 for *P. sojae* (Ma *et al.*, 2017).

The pathogen, however, protects *Ps*XEG1 from host inhibition by secreting a more tightly binding, truncated paralog of *Ps*XEG1, called *Ps*XLP1. *Ps*XLP1 has lost one of the residues critical for enzyme activity and has no known function in pathogenicity, other than to intercept *Gm*GIP1 (Ma *et al.*, 2017). This leads to a 'bodyguard' model in which *Ps*XLP1 has evolved as a paralogous decoy to neutralize *Gm*GIP1 and prevent inhibition of *Ps*XEG1 (**Figure 1A**). The conservation of PsXLP1 paralogs across *Phytophthora* species suggests that these oomycete pathogens use these paralogous decoys frequently to bodyguard *Ps*XEG1-like effectors (Ma *et al.*, 2017).

The publication of Ma *et al.*, 2017 (Ma *et al.*, 2017) strengthens this exciting new area in the field of host-pathogen interactions. The described pathogen decoy is distinct from the decoy function of host-derived factors as described in the Decoy Model, where decoys act as co-receptors or sensors with immune receptors (van der Hoorn and Kamoun, 2008). Both host- and pathogen-derived decoys are specialized, often paralogous forms of a functional effector target or effector, respectively. But in contrast to host derived decoys that trap effectors to mount effector-triggered immunity (ETI), pathogen decoys protect virulence factors as bodyguards.

Two recent publications describe a distinct but similar bodyguard strategy used by the gram-negative bacterium *Xanthomonas oryzae* (Ji *et al.*, 2016; Read *et al.*, 2016). *X. oryzae* pathovars *oryzae* (*Xoo*) and *oryzicola* (*Xoc*) cause devastating bacterial blight (BB) and bacterial leaf streak (BLS) in rice (*Oryza sativa*), respectively. *Xanthomonas* strains express a variety of Type-III effectors, including many Transcription activator-like effectors (TALEs) (Kay and Bonas, 2009; Boch and Bonas, 2010). TALEs are major virulence determinants that act by trans-activating host genes in the plant cell nucleus by binding to promoter elements (Effector-binding Elements, EBEs) in a sequence-specific manner.

TALEs contain several nuclear localization signals (NLSs) that target these effectors to the host cell nucleus. TALEs also carry a C-terminal acidic activation domain (AD) that activates host gene expression, and a central repeat domain that directly binds to the EBE in the DNA. Resistance to *Xoo* has mostly been attributed to the action of TALEs, either by polymorphisms in EBEs that prevent the induction of susceptibility (*S*) genes, or by the induction of executor genes that carry EBEs embedded in their promoter and induce immune responses as dominant *R* genes (Hutin *et al.*, 2016).

However, some TALEs are recognized by classical *R* genes that encode NLRs (Nod-like Receptors). The TALE AvrBs4 of *X. euvesicatoria* is recognized by the Bs4-encoded NLR protein in tomato (Schornack *et al.*, 2004), whilst all tested, full length TALEs of *Xoo* are recognized by the *Xa1*-encoded NLR in rice (Ji *et al.*, 2016). Interestingly, however, both the Yang (Iowa State University, USA) and Bogdanove (Cornell University, USA) laboratories found a new class of truncated TALEs that could suppress TALE recognition by NLRs.

The Yang group demonstrated that TALE perception by the NLR Xa1 is suppressed by two groups of truncated TALE variants that interfere with Xa1, which they coined “iTALEs”. The frequent occurrence of iTALEs could explain the narrow resistance spectrum of Xa1, even though it appears to perceive most, if not all TALEs (Ji *et al.*, 2016). Similarly, the Bogdanove group found that *Xo1* resistance in a different rice variety, which is also mediated by the perception of TALEs (Triplett *et al.*, 2016), can be blocked by truncated TALEs – “truncTALEs” - such as Tal2h from *Xoc* strain BLS256. In contrast to *Xa1*, *Xo1* has not been cloned yet, but the *Xo1* locus is allelic to *Xa1*, confers resistance to both *Xoc* and *Xoo*, and contains six candidate *R* genes encoding NLRs, indicating that *Xo1* encodes an NLR conferring TALE perception (Ji *et al.*, 2016; Kawahara *et al.*, 2013; Triplett *et al.*, 2016).

The truncated Tal2h TALE lacks N- and C-terminal regions that are important for DNA binding and this implies that Tal2h does not act by binding host DNA (Read *et al.*, 2016). Both iTALEs and truncTALEs were previously annotated as pseudogenes due to their N- and C-terminal deletions. Both studies conclude that the truncated TALE proteins bind to TALE-recognizing NLR proteins to block the TALE binding site without activating the NLR receptor, to prevent the recognition of full length TALEs that act as virulence factors (**Figure 1B**). Even though there is strong genetic evidence, this hypothesis still requires experimental support to demonstrate a direct interaction of TALEs with the Xa1/Xo1 NLRs and that truncated TALEs interfere with this interaction by having a higher affinity for NLRs. These are notoriously challenging experiments as NLR proteins are difficult to produce and purify.

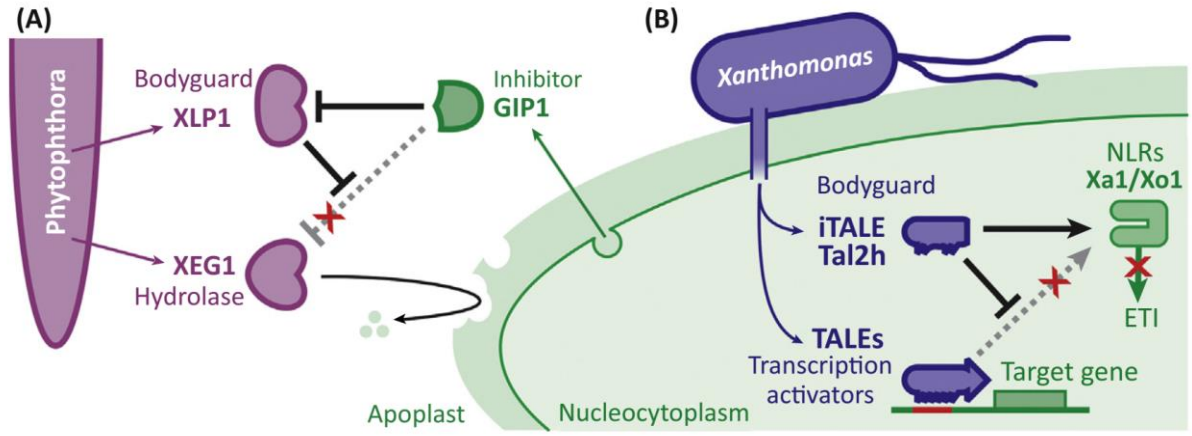


Figure 1. Model of Two Bodyguards at Work. **(A)** Effector decoy XLP1 prevents host protein GIP1 from inhibiting glucanase XEG1. **(B)** Truncated TALEs prevent TALE effectors from being recognized by immune receptors Xa1/Xo1. Abbreviations: ETI, effector-triggered immunity; GIP1, glucanase inhibitor protein; TALE, Transcription activator-like effectors; Xa1, *R* gene in rice; XEG1, Xyloglucanase; XLP1, XEG1-Like Protein; Xo1, *R* gene in rice.

Both XLP1 and truncated TALEs illustrate that paralogous decoys can evolve to protect important virulence factors to prevent their inactivation or recognition. Theoretically, these bodyguards have no other function in disease development, meaning that they have no role in the absence of the virulence factor they are mimicking, or in the absence of host proteins that

inactivate or recognize this virulence factor. These bodyguard effectors probably have a higher affinity to the host protein and/or higher abundance when compared to the corresponding virulence factor. Both mechanisms allow the virulence factor to remain unaffected or undetected.

The evolution of a bodyguard that protects a single virulence factor may seem very costly. From an evolutionary standpoint, it might seem more efficient for a pathogen to develop effectors that interfere broadly with immune signaling. However, this bodyguard strategy might be worthwhile for critical effectors. In addition, the example of truncated TALEs illustrates how one bodyguard can simultaneously protect a larger class of virulence factors.

The production of decoy effector proteins by two unrelated pathogens highlights a common strategy of pathogens to protect virulence factors. These observations encourage new searches in pathogen genomes to investigate truncated or seemingly inactive effector proteins that were previously considered pseudogenes. Increased knowledge on these mechanisms will enable us to engineer the host immune system to circumvent manipulation by bodyguard effectors.

Acknowledgements. The authors are financially supported by ECR Consolidator grant 616449 'GreenProteases', the University of Oxford and The Clarendon Fund.

Author contributions

JK identified the connection between the discussed papers. RALvdH generated the figure with feedback from JKP and JK. JKP wrote the first draft of the paper. JKP and JK edited the manuscript with feedback from RALvdH.

References

Ade, J., DeYoung, B.J., Golstein, C. and Innes, R.W. (2007) Indirect activation of a plant nucleotide binding site–leucine-rich repeat protein by a bacterial protease. *Proc. Natl. Acad. Sci.*, **104**, 2531–2536.

Albert, I., Böhm, H., Albert, M., et al. (2015) An RLP23–SOBIR1–BAK1 complex mediates NLP-triggered immunity. *Nat. Plants*, **1**, 15140.

Ashfield, T., Ong, L.E., Nobuta, K., Schneider, C.M. and Innes, R.W. (2004) Convergent evolution of disease resistance gene specificity in two flowering plant families. *Plant Cell*, **16**, 309–318.

Ashikawa, I., Hayashi, N., Yamane, H., Kanamori, H., Wu, J., Matsumoto, T., Ono, K. and Yano, M. (2008) Two adjacent nucleotide-binding site–leucine-rich repeat class genes are required to confer *Pikm*-specific rice blast resistance. *Genetics*, **180**, 2267–2276.

Axtell, M.J. and Staskawicz, B.J. (2003) Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4. *Cell*, **112**, 369–377.

Bailey, P.C., Schudoma, C., Jackson, W., Baggs, E., Dagdas, G., Haerty, W., Moscou, M. and Krasileva, K.V. (2018) Dominant integration locus drives continuous diversification of plant immune receptors with exogenous domain fusions. *Genome Biol.*, **19**, 23.

Balagué, C., Gouget, A., Bouchez, O., Souriac, C., Haget, N., Boutet-Mercey, S., Govers, F., Roby, D. and Canut, H. (2017) The *Arabidopsis thaliana* lectin receptor kinase LecRK-I.9 is required for full resistance to *Pseudomonas syringae* and affects jasmonate signalling. *Mol. Plant Pathol.*, **18**, 937–948.

Bally, J., Nakasugi, K., Jia, F., et al. (2015) The extremophile *Nicotiana benthamiana* has traded viral defence for early vigour. *Nat. Plants*, **1**, 15165.

Balmuth, A. and Rathjen, J.P. (2007) Genetic and molecular requirements for function of the Pto/Prf effector recognition complex in tomato and *Nicotiana benthamiana*. *Plant J.*, **51**, 978–990.

Baudin, M., Hassan, J.A., Schreiber, K.J. and Lewis, J.D. (2017) Analysis of the ZAR1 immune complex reveals determinants for immunity and molecular interactions. *Plant Physiol.*, **174**, 2038–2053.

Beck, S., Michalski, A., Raether, O., et al. (2015) The impact II, a very high-resolution quadrupole time-of-flight instrument (QTOF) for deep shotgun proteomics. *Mol. Cell. Proteomics*, **14**, 2014–2029.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S. and Nekrasov, V. (2013) Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods*, **9**, 39.

Benedetti, M., Pontiggia, D., Raggi, S., Cheng, Z., Scaloni, F., Ferrari, S., Ausubel, F.M., Cervone, F. and De Lorenzo, G. (2015) Plant immunity triggered by engineered in vivo release of oligogalacturonides, damage-associated molecular patterns. *Proc. Natl. Acad. Sci.*, **112**, 5533–5538.

Bent, A.F., Kunkel, B.N., Dahlbeck, D., Brown, K.L., Schmidt, R., Giraudat, J., Leung, J. and Staskawicz, B.J. (1994) RPS2 of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science*, **265**, 1856–1860.

Bernoux, M., Burdett, H., Williams, S.J., et al. (2016) Comparative analysis of the flax immune receptors L6 and L7 suggests an equilibrium-based switch activation model. *Plant Cell*, **28**, 146–159.

Biniossek, M.L., Niemer, M., Maksimchuk, K., et al. (2016) Identification of protease specificity by combining proteome-derived peptide libraries and quantitative proteomics. *Mol. Cell. Proteomics*, **15**, 2515–2524.

Boch, J. and Bonas, U. (2010) *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.*, **48**, 419–436.

Böhm, H., Albert, I., Oome, S., Raaymakers, T.M., Van den Ackerveken, G.F.J.M. and Nürnberger, T. (2014) A conserved peptide pattern from a widespread microbial virulence factor triggers pattern-induced immunity in Arabidopsis. *PLOS Pathog.*, **10**, e1004491.

Bombarely, A., Moser, M., Amrad, A., et al. (2016) Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants*, **2**, 16074.

Bombarely, A., Rosli, H.G., Vrebalov, J., Moffett, P., Mueller, L.A. and Martin, G.B. (2012) A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Mol. Plant. Microbe Interact.*, **25**, 1523–1530.

Bonardi, V., Tang, S., Stallmann, A., Roberts, M., Cherkis, K. and Dangl, J.L. (2011) Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proc. Natl. Acad. Sci.*, **108**, 16463–16468.

Bouwmeester, K., Han, M., Blanco-Portales, R., Song, W., Weide, R., Guo, L.-Y., Vossen, E.A.G. van der and Govers, F. (2014) The Arabidopsis lectin receptor kinase LecRK-I.9 enhances resistance to *Phytophthora infestans* in Solanaceous plants. *Plant Biotechnol. J.*, **12**, 10–16.

Bouwmeester, K., Sain, M. de, Weide, R., Gouget, A., Klamer, S., Canut, H. and Govers, F. (2011) The lectin receptor kinase LecRK-I.9 is a novel *Phytophthora* resistance component and a potential host target for a RXLR effector. *PLOS Pathog.*, **7**, e1001327.

Boyle, E.A., Li, Y.I. and Pritchard, J.K. (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell*, **169**, 1177–1186.

Brutus, A., Sicilia, F., Macone, A., Cervone, F. and De Lorenzo, G. (2010) A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. *Proc. Natl. Acad. Sci.*, **107**, 9452–9457.

Büschges, R., Hollricher, K., Panstruga, R., et al. (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell*, **88**, 695–705.

Butterbach, P., Verlaan, M.G., Dullemans, A., Lohuis, D., Visser, R.G.F., Bai, Y. and Kormelink, R. (2014) Tomato yellow leaf curl virus resistance by *Ty-1* involves increased cytosine methylation of viral genomes and is compromised by cucumber mosaic virus infection. *Proc. Natl. Acad. Sci.*, **111**, 12942–12947.

Cai, D., Kleine, M., Kifle, S., et al. (1997) Positional cloning of a gene for nematode resistance in sugar beet. *Science*, **275**, 832–834.

Caicedo, A.L. and Schaal, B.A. (2004) Heterogeneous evolutionary processes affect *R* gene diversity in natural populations of *Solanum pimpinellifolium*. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 17444–17449.

Caldwell, K.S. and Michelmore, R.W. (2009) *Arabidopsis thaliana* genes encoding defense signaling and recognition proteins exhibit contrasting evolutionary dynamics. *Genetics*, **181**, 671–684.

Caplan, J.L., Mamillapalli, P., Burch-Smith, T.M., Czymmek, K. and Dinesh-Kumar, S.P. (2008) Chloroplastic protein NRIP1 mediates innate immune receptor recognition of a viral effector. *Cell*, **132**, 449–462.

Carbonell, A., Takeda, A., Fahlgren, N., Johnson, S.C., Cuperus, J.T. and Carrington, J.C. (2014) New generation of artificial microRNA and synthetic trans-acting small interfering RNA vectors for efficient gene silencing in Arabidopsis. *Plant Physiol.*, **165**, 15–29.

Carter, M.E., Helm, M., Chapman, A., Wan, E., Restrepo Sierra, A.M., Innes, R., Bogdanove, A.J. and Wise, R.P. (2018) Convergent evolution of effector protease recognition by Arabidopsis and barley. *Mol. Plant. Microbe Interact.*

Casimiro-Soriguer, C.S., Muñoz-Mérida, A. and Pérez-Pulido, A.J. (2017) Sma3s: A universal tool for easy functional annotation of proteomes and transcriptomes. *Proteomics*, **17**.

Catanzariti, A.-M., Lim, G.T.T. and Jones, D.A. (2015) The tomato *I-3* gene: a novel gene for resistance to Fusarium wilt disease. *New Phytol.*, **207**, 106–118.

Cesari, S., Thilliez, G., Ribot, C., et al. (2013) The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. *Plant Cell*, **25**, 1463–1481.

Chakrabarti, A., Panter, S.N., Harrison, K., Jones, J.D.G. and Jones, D.A. (2009) Regions of the Cf-9B disease resistance protein able to cause spontaneous necrosis in *Nicotiana benthamiana* lie within the region controlling pathogen recognition in tomato. *Mol. Plant. Microbe Interact.*, **22**, 1214–1226.

Chen, C., Chen, H., Lin, Y.-S., et al. (2014) A two-locus interaction causes interspecific hybrid weakness in rice. *Nat. Commun.*, **5**, 3357.

Cheng, C.-Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S. and Town, C.D. (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.*, **89**, 789–804.

Cheng, W., Munkvold, K.R., Gao, H., et al. (2011) Structural analysis of *Pseudomonas syringae* AvrPtoB bound to host BAK1 reveals two similar kinase-interacting domains in a type III effector. *Cell Host Microbe*, **10**, 616–626.

Cheng, Z., Li, J.-F., Niu, Y., et al. (2015) Pathogen-secreted proteases activate a novel plant immune pathway. *Nature*, **521**, 213–216.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nürnberger, T., Jones, J.D.G., Felix, G. and Boller, T. (2007) A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature*, **448**, 497–500.

Choi, J., Tanaka, K., Cao, Y., Qi, Y., Qiu, J., Liang, Y., Lee, S.Y. and Stacey, G. (2014) Identification of a plant receptor for extracellular ATP. *Science*, **343**, 290–294.

Chu, Z., Yuan, M., Yao, J., et al. (2006) Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes Dev.*, **20**, 1250–1255.

Clarkson, J.J., Dodsworth, S. and Chase, M.W. (2017) Time-calibrated phylogenetic trees establish a lag between polyploidisation and diversification in *Nicotiana* (Solanaceae). *Plant Syst. Evol.*, **303**, 1001–1012.

Clemente, T. (2006) *Nicotiana* (*Nicotiana tobaccum*, *Nicotiana benthamiana*). In K. Wang, ed. *Agrobacterium Protocols*. Methods in Molecular Biology (Clifton, N.J.). Humana Press, pp. 143–154.

Coaker, G., Falick, A. and Staskawicz, B.J. (2005) Activation of a phytopathogenic bacterial effector protein by a eukaryotic cyclophilin. *Science*, **308**, 548–550.

Consortium, T.P.G.S. (2011) Genome sequence and analysis of the tuber crop potato. *Nature*, **475**, 189–195.

Consortium, T.T.G. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N. and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics*, **13**, 2513–2526.

Cox, J. and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V. and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.*, **10**, 1794–1805.

Cui, H., Tsuda, K. and Parker, J.E. (2015) Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. *Annu. Rev. Plant Biol.*, **66**, 487–511.

Day, B., Dahlbeck, D., Huang, J., Chisholm, S.T., Li, D. and Staskawicz, B.J. (2005) Molecular basis for the RIN4 negative regulation of *RPS2* disease resistance. *Plant Cell*, **17**, 1292–1305.

Debieu, M., Huard-Chauveau, C., Genissel, A., Roux, F. and Roby, D. (2016) Quantitative disease resistance to the bacterial pathogen *Xanthomonas campestris* involves an Arabidopsis immune receptor pair and a gene of unknown function. *Mol. Plant Pathol.*, **17**, 510–520.

DeYoung, B.J., Qi, D., Kim, S.-H., Burke, T.P. and Innes, R.W. (2012) Activation of a plant nucleotide binding-leucine rich repeat disease resistance protein by a modified self protein. *Cell. Microbiol.*, **14**, 1071–1084.

Deyoung, B.J., Qi, D., Kim, S.-H., Burke, T.P. and Innes, R.W. (2012) Activation of a plant nucleotide binding-leucine rich repeat disease resistance protein by a modified self protein. *Cell. Microbiol.*, **14**, 1071–1084.

Dixon, M.S., Golstein, C., Thomas, C.M., Biezen, E.A. van der and Jones, J.D.G. (2000) Genetic complexity of pathogen perception by plants: the example of *Rcr3*, a tomato gene required specifically by *Cf-2*. *Proc. Natl. Acad. Sci.*, **97**, 8807–8814.

Dixon, M.S., Hatzixanthis, K., Jones, D.A., Harrison, K. and Jones, J.D.G. (1998) The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. *Plant Cell Online*, **10**, 1915–1925.

Dixon, M.S., Jones, D.A., Keddie, J.S., Thomas, C.M., Harrison, K. and Jones, J.D.G. (1996) The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell*, **84**, 451–459.

Dobzhansky, T. (1973) Nothing in Biology Makes Sense except in the Light of Evolution. *Am. Biol. Teach.*, **35**, 125–129.

Dodds, P.N., Lawrence, G.J., Catanzariti, A.-M., Ayliffe, M.A. and Ellis, J.G. (2004) The *Melampsora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell*, **16**, 755–768.

Dodds, P.N., Lawrence, G.J., Catanzariti, A.-M., Teh, T., Wang, C.-I.A., Ayliffe, M.A., Kobe, B. and Ellis, J.G. (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc. Natl. Acad. Sci.*, **103**, 8888–8893.

Doehlemann, G. and Hemetsberger, C. (2013) Apoplastic immunity and its suppression by filamentous plant pathogens. *New Phytol.*, **198**, 1001–1016.

Doench, J.G., Fusi, N., Sullender, M., et al. (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 184–191.

Dong, S., Stam, R., Cano, L.M., et al. (2014) Effector specialization in a lineage of the irish potato famine pathogen. *Science*, **343**, 552–555.

Ellis, J.G., Lagudah, E.S., Spielmeyer, W. and Dodds, P.N. (2014) The past, present and future of breeding rust resistant wheat. *Front. Plant Sci.*, **5**.

Engineer, C.B., Ghassemian, M., Anderson, J.C., Peck, S.C., Hu, H. and Schroeder, J.I. (2014) Carbonic anhydrases, *EPF2* and a novel protease mediate CO_2 control of stomatal development. *Nature*, **513**, 246–250.

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.-M., Werner, S., Jones, J.D.G., Patron, N.J. and Marillonnet, S. (2014) A Golden Gate modular cloning toolbox for plants. *ACS Synth. Biol.*, **3**, 839–843.

Esse, H.P. van, Klooster, J.W. van't, Bolton, M.D., Yadeta, K.A., Baarlen, P. van, Boeren, S., Vervoort, J., Wit, P.J.G.M. de and Thomma, B.P.H.J. (2008) The *Cladosporium fulvum* virulence protein Avr2 inhibits host proteases required for basal defense. *Plant Cell Online*, **20**, 1948–1963.

Felix, G., Duran, J.D., Volko, S. and Boller, T. (1999) Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J. Cell Mol. Biol.*, **18**, 265–276.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evol. Int. J. Org. Evol.*, **39**, 783–791.

Fernandez-Pozo, N., Rosli, H.G., Martin, G.B. and Mueller, L.A. (2015) The SGN VIGS Tool: user-friendly software to design Virus-Induced Gene Silencing (VIGS) constructs for functional genomics. *Mol. Plant*, **8**, 486–488.

Finn, R.D., Coggill, P., Eberhardt, R.Y., et al. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.

Fradin, E.F., Zhang, Z., Ayala, J.C.J., Castroverde, C.D.M., Nazar, R.N., Robb, J., Liu, C.-M. and Thomma, B.P.H.J. (2009) Genetic dissection of Verticillium wilt resistance mediated by tomato *Ve1*. *Plant Physiol.*, **150**, 320–332.

Fu, D., Uauy, C., Distelfeld, A., Blechl, A., Epstein, L., Chen, X., Sela, H., Fahima, T. and Dubcovsky, J. (2009) A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. *Science*, **323**, 1357–1360.

Fu, L., Niu, B., Zhu, Z., Wu, S. and Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**, 3150–3152.

Fujisaki, K., Abe, Y., Kanzaki, E., et al. (2017) An unconventional NOI/RIN4 domain of a rice NLR protein binds host EXO70 protein to confer fungal immunity. *bioRxiv*, 239400.

Fukuoka, S., Saka, N., Koga, H., et al. (2009) Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science*, **325**, 998–1001.

Gabriëls, S.H.E.J., Vossen, J.H., Ekengren, S.K., et al. (2007) An NB-LRR protein required for HR signalling mediated by both extra- and intracellular resistance proteins. *Plant J.*, **50**, 14–28.

Gao, Z., Chung, E.-H., Eitas, T.K. and Dangl, J.L. (2011) Plant intracellular innate immune receptor Resistance to Pseudomonas syringae pv. maculicola 1 (RPM1) is activated at, and functions on, the plasma membrane. *Proc. Natl. Acad. Sci.*, **108**, 7619–7624.

Ghorbani, S., Hoogewijs, K., Pečenková, T., et al. (2016) The SBT6.1 subtilase processes the GOLVEN1 peptide controlling cell elongation. *J. Exp. Bot.*, **67**, 4877–4887.

Giannakopoulou, A., Steele, J.F.C., Segretin, M.E., Bozkurt, T.O., Zhou, J., Robatzek, S., Banfield, M.J., Pais, M. and Kamoun, S. (2015) Tomato I2 immune receptor can be engineered to confer partial resistance to the oomycete *Phytophthora infestans* in addition to the fungus *Fusarium oxysporum*. *Mol. Plant. Microbe Interact.*, **28**, 1316–1329.

Gómez-Gómez, L. and Boller, T. (2000) FLS2: an LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Mol. Cell*, **5**, 1003–1011.

Goodin, M.M., Zaitlin, D., Naidu, R.A. and Lommel, S.A. (2008) *Nicotiana benthamiana*: its history and future as a model for plant–pathogen interactions. *Mol. Plant. Microbe Interact.*, **21**, 1015–1026.

Goodstein, D.M., Shu, S., Howson, R., et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.*, **40**, D1178–D1186.

Gou, J.-Y., Li, K., Wu, K., et al. (2015) Wheat stripe rust resistance protein WKS1 reduces the ability of the thylakoid-associated ascorbate peroxidase to detoxify reactive oxygen species. *Plant Cell*, **27**, 1755–1770.

Grosse-Holz, F., Madeira, L., Zahid, M.A., et al. (2018) Three unrelated protease inhibitors enhance accumulation of pharmaceutical recombinant proteins in *Nicotiana benthamiana*. *Plant Biotechnol. J.*, **16**, 1797–1810.

Grosse-Holz, F.M., Kelly, S., Blaskowski, S., Kaschani, F., Kaiser, M. and Hoorn, R.A.L. van der (2018) The transcriptome, extracellular proteome and active secretome of agroinfiltrated *Nicotiana benthamiana* uncover a large, diverse protease repertoire. *Plant Biotechnol. J.*, **16**, 1068–1084.

Gu, K., Tian, D., Qiu, C. and Yin, Z. (2009) Transcription activator-like type III effector AvrXa27 depends on OsTFIIAgamma5 for the activation of Xa27 transcription in rice that triggers disease resistance to Xanthomonas oryzae pv. oryzae. *Mol. Plant Pathol.*, **10**, 829–835.

Gu, K., Yang, B., Tian, D., et al. (2005) *R* gene expression induced by a type-III effector triggers disease resistance in rice. *Nature*, **435**, 1122–1125.

Guillen, K. de, Ortiz-Vallejo, D., Gracy, J., Fournier, E., Kroj, T. and Padilla, A. (2015) Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi. *PLOS Pathog.*, **11**, e1005228.

Gumtow, R., Wu, D., Uchida, J. and Tian, M. (2017) A *Phytophthora palmivora* extracellular Cystatin-like protease inhibitor targets Papain to contribute to virulence on papaya. *Mol. Plant. Microbe Interact.*, **31**, 363–373.

Haas, B.J., Papanicolaou, A., Yassour, M., et al. (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.*, **8**, 1494–1512.

Herrera-Foessel, S.A., Singh, R.P., Lillemo, M., Huerta-Espino, J., Bhavani, S., Singh, S., Lan, C., Calvo-Salazar, V. and Lagudah, E.S. (2014) *Lr67/Yr46* confers adult plant resistance to stem rust and powdery mildew in wheat. *Theor. Appl. Genet.*, **127**, 781–789.

Hind, S.R., Strickler, S.R., Boyle, P.C., et al. (2016) Tomato receptor FLAGELLIN-SENSING 3 binds flgII-28 and activates the plant immune system. *Nat. Plants*, **2**, 16128.

Hirakawa, H., Shirasawa, K., Miyatake, K., et al. (2014) Draft genome sequence of eggplant (*Solanum melongena* L.): the representative *Solanum* species indigenous to the old world. *DNA Res.*, **21**, 649–660.

Hogenhout, S.A., Hoorn, R.A.L. van der, Terauchi, R. and Kamoun, S. (2009) Emerging concepts in effector biology of plant-associated organisms. *Mol. Plant. Microbe Interact.*, **22**, 115–122.

van der Hoorn, R.A.L. and Kamoun, S. (2008) From guard to decoy: a new model for perception of plant pathogen effectors. *Plant Cell Online*, **20**, 2009–2017.

van der Hoorn, R.A.L., Kruijt, M., Roth, R., Brandwagt, B.F., Joosten, M.H.A.J. and de Wit, P.J.G.M. (2001) Intragenic recombination generated two distinct *Cf* genes that mediate AVR9 recognition in the natural population of *Lycopersicon pimpinellifolium*. *Proc. Natl. Acad. Sci.*, **98**, 10493–10498.

van der Hoorn, R.A.L., Roth, R. and de Wit, P.J.G.M. (2001) Identification of distinct specificity determinants in resistance protein Cf-4 allows construction of a Cf-9 mutant that confers recognition of avirulence protein AVR4. *Plant Cell Online*, **13**, 273–285.

Hörger, A.C., Ilyas, M., Stephan, W., Tellier, A., van der Hoorn, R.A.L. and Rose, L.E. (2012) Balancing selection at the tomato *Rcr3* guardee gene family maintains variation in strength of pathogen defense. *PLoS Genet*, **8**, e1002813.

Hou, S., Wang, X., Chen, D., Yang, X., Wang, M., Turrà, D., Pietro, A.D. and Zhang, W. (2014) The Secreted Peptide PIP1 Amplifies Immunity through Receptor-Like Kinase 7. *PLOS Pathog.*, **10**, e1004331.

Huang, S., Antony, G., Li, T., Liu, B., Obasa, K., Yang, B. and White, F.F. (2016) The broadly effective recessive resistance gene *xa5* of rice is a virulence effector-dependent quantitative trait for bacterial blight. *Plant J. Cell Mol. Biol.*, **86**, 186–194.

Hückelhoven, R. and Panstruga, R. (2011) Cell biology of the plant–powdery mildew interaction. *Curr. Opin. Plant Biol.*, **14**, 738–746.

Huffaker, A., Pearce, G. and Ryan, C.A. (2006) An endogenous peptide signal in *Arabidopsis* activates components of the innate immune response. *Proc. Natl. Acad. Sci.*, **103**, 10098–10103.

Hummel, A.W., Doyle, E.L. and Bogdanove, A.J. (2012) Addition of transcription activator-like effector binding sites to a pathogen strain-specific rice bacterial blight resistance gene makes it effective against additional strains and against bacterial leaf streak. *New Phytol.*, **195**, 883–893.

Humphry, M., Bednarek, P., Kemmerling, B., et al. (2010) A regulon conserved in monocot and dicot plants defines a functional module in antifungal plant immunity. *Proc. Natl. Acad. Sci.*, **107**, 21896–21901.

Humphry, M., Consonni, C. and Panstruga, R. (2006) *mlo*-based powdery mildew immunity: silver bullet or simply non-host resistance? *Mol. Plant Pathol.*, **7**, 605–610.

Hurlburt, N.K., Chen, L.-H., Stergiopoulos, I. and Fisher, A.J. (2018) Structure of the *Cladosporium fulvum* Avr4 effector in complex with $(GlcNAc)_6$ reveals the ligand-binding mechanism and uncouples its intrinsic function from recognition by the Cf-4 resistance protein. *PLOS Pathog.*, **14**, e1007263.

Hutin, M., Césari, S., Chalvon, V., Michel, C., Tran, T.T., Boch, J., Koebnik, R., Szurek, B. and Kroj, T. (2016) Ectopic activation of the rice NLR heteropair RGA4/RGA5 confers resistance to bacterial blight and bacterial leaf streak diseases. *Plant J. Cell Mol. Biol.*, **88**, 43–55.

Ilyas, M., Hörger, A.C., Bozkurt, T.O., et al. (2015) Functional divergence of two secreted immune proteases of tomato. *Curr. Biol.*, **25**, 2300–2306.

Iorizzo, M., Ellison, S., Senalik, D., et al. (2016) A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.*, **48**, 657–666.

Ishibashi, K., Kezuka, Y., Kobayashi, C., Kato, M., Inoue, T., Nonaka, T., Ishikawa, M., Matsumura, H. and Katoh, E. (2014) Structural basis for the recognition–evasion arms race between Tomato mosaic virus and the resistance gene *Tm-1*. *Proc. Natl. Acad. Sci.*, **111**, E3486–E3495.

Iyer, A.S. and McCouch, S.R. (2004) The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. *Mol. Plant. Microbe Interact.*, **17**, 1348–1354.

Iyer-Pascuzzi, A.S., Jiang, H., Huang, L. and McCouch, S.R. (2008) Genetic and functional characterization of the rice bacterial blight disease resistance gene *xa5*. *Phytopathology*, **98**, 289–295.

Ji, Z., Ji, C., Liu, B., Zou, L., Chen, G. and Yang, B. (2016) Interfering TAL effectors of *Xanthomonas oryzae* neutralize *R*-gene-mediated plant disease resistance. *Nat. Commun.*, **7**, 13435.

Johal, G.S. and Briggs, S.P. (1992) Reductase activity encoded by the *HM1* disease resistance gene in maize. *Science*, **258**, 985–987.

Jones, D.A., Thomas, C.M., Hammond-Kosack, K.E., Balint-Kurti, P.J. and Jones, J.D.G. (1994) Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science*, **266**, 789–793.

Jones, J.D.G. and Dangl, J.L. (2006) The plant immune system. *Nature*, **444**, 323–329.

Jonge, R. de, Esse, H.P. van, Maruthachalam, K., et al. (2012) Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc. Natl. Acad. Sci.*, **109**, 5110–5115.

Joosten, M.H.A.J., Cozijnsen, T.J. and Wit, P.J.G.M. de (1994) Host resistance to a fungal tomato pathogen lost by a single base-pair change in an avirulence gene. *Nature*, **367**, 384–386.

van Kan, J.A., van den Ackerveken, G.F.J.M. and de Wit, P.J.G.M. (1991) Cloning and characterization of cDNA of avirulence gene *avr9* of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. *Mol. Plant-Microbe Interact. MPMI*, **4**, 52–59.

Kang, B.-C., Yeam, I. and Jahn, M.M. (2005) Genetics of plant virus resistance. *Annu. Rev. Phytopathol.*, **43**, 581–621.

Kaschani, F., Shabab, M., Bozkurt, T., et al. (2010) An effector-targeted protease contributes to defense against *Phytophthora infestans* and is under diversifying selection in natural hosts. *Plant Physiol.*, **154**, 1794–1804.

Katsuragi, Y., Takai, R., Furukawa, T., Hirai, H., Morimoto, T., Katayama, T., Murakami, T. and Che, F.-S. (2015) CD2-1, the C-terminal region of flagellin, modulates the induction of immune responses in rice. *Mol. Plant. Microbe Interact.*, **28**, 648–658.

Kawahara, Y., Bastide, M. de la, Hamilton, J.P., et al. (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, **6**, 4.

Kay, S. and Bonas, U. (2009) How *Xanthomonas* type III effectors manipulate the host plant. *Curr. Opin. Microbiol.*, **12**, 37–43.

Keller, O., Kollmar, M., Stanke, M. and Waack, S. (2011) A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics*, **27**, 757–763.

Keller, O., Odronitz, F., Stanke, M., Kollmar, M. and Waack, S. (2008) Scipio: using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. *BMC Bioinformatics*, **9**, 278.

Kessens, R., Ashfield, T., Kim, S.H. and Innes, R.W. (2014) Determining the GmRIN4 requirements of the soybean disease resistance proteins Rpg1b and Rpg1r using a *Nicotiana glutinosa*-based agroinfiltration system. *PLOS ONE*, **9**, e108159.

Kim, H.-S., Desveaux, D., Singer, A.U., Patel, P., Sondek, J. and Dangl, J.L. (2005) The *Pseudomonas syringae* effector AvrRpt2 cleaves its C-terminally acylated target, RIN4, from Arabidopsis membranes to block RPM1 activation. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 6496–6501.

Kim, M.G., Geng, X., Lee, S.Y. and Mackey, D. (2009) The *Pseudomonas syringae* type III effector AvrRpm1 induces significant defenses by activating the Arabidopsis nucleotide-binding leucine-rich repeat protein RPS2. *Plant J.*, **57**, 645–653.

Kim, S., Park, M., Yeom, S.-I., et al. (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.*, **46**, 270–278.

Kim, S.H., Qi, D., Ashfield, T., Helm, M. and Innes, R.W. (2016) Using decoys to expand the recognition specificity of a plant disease resistance protein. *Science*, **351**, 684–687.

Kim, Y.J., Lin, N.-C. and Martin, G.B. (2002) Two distinct *Pseudomonas* effector proteins interact with the Pto kinase and activate plant immunity. *Cell*, **109**, 589–598.

van't Klooster, J.W., van der Kamp, M.W., Vervoort, J., Beekwilder, J., Boeren, S., Joosten, M.H.A.J., Thomma, B.P.H.J. and de Wit, P.J.G.M. (2011) Affinity of Avr2 for tomato cysteine protease Rcr3 correlates with the Avr2-triggered *Cf-2*-mediated hypersensitive response. *Mol. Plant Pathol.*, **12**, 21–30.

Klymiuk, V., Yaniv, E., Huang, L., et al. (2018) Cloning of the wheat *Yr15* resistance gene sheds light on the plant tandem kinase-pseudokinase family. *Nat. Commun.*, **9**, 3735.

Kooman-Gersmann, M., Honee, G., Bonnema, G. and de Wit, P.J.G.M. (1996) A high-affinity binding site for the AVR9 peptide elicitor of *Cladosporium fulvum* is present on plasma membranes of tomato and other Solanaceous plants. *Plant Cell*, **8**, 929–938.

Kooman-Gersmann, M., Vogelsang, R., Vossen, P., van den Hooven, H.W., Mahé, E., Honée, G. and de Wit, P.J.G.M. (1998) Correlation between binding affinity and necrosis-inducing activity of mutant AVR9 peptide elicitors. *Plant Physiol.*, **117**, 609–618.

Kourelis, J. and van der Hoorn, R.A.L. (2018) Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *Plant Cell*, **30**, 285–299.

Kourelis, J., Hoorn, van der R.A.L. and Sueldo, D.J. (2016) Decoy engineering: the next step in resistance breeding. *Trends Plant Sci.*, **21**, 371–373.

Kourelis, J., Kaschani, F., Grosse-Holz, F.M., Homma, F., Kaiser, M. and van der Hoorn, R.A.L. (2018) Re-annotated *Nicotiana benthamiana* gene models for enhanced proteomics and reverse genetics. *bioRxiv*, 373506.

Krasileva, K.V., Dahlbeck, D. and Staskawicz, B.J. (2010) Activation of an Arabidopsis resistance protein is specified by the *in planta* association of its leucine-rich repeat domain with the cognate oomycete effector. *Plant Cell*, **22**, 2444–2458.

Krattinger, S.G., Lagudah, E.S., Spielmeyer, W., Singh, R.P., Huerta-Espino, J., McFadden, H., Bossolini, E., Selter, L.L. and Keller, B. (2009) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science*, **323**, 1360–1363.

Kroj, T., Chanclud, E., Michel-Romiti, C., Grand, X. and Morel, J.-B. (2016) Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytol.*, **210**, 618–626.

Krüger, J., Thomas, C.M., Golstein, C., Dixon, M.S., Smoker, M., Tang, S., Mulder, L. and Jones, J.D.G. (2002) A tomato cysteine protease required for *Cf-2*-dependent disease resistance and suppression of autonecrosis. *Science*, **296**, 744–747.

Kruijt, M., Kip, D.J., Joosten, M.H.A.J., Brandwagt, B.F. and de Wit, P.J.G.M. (2005) The *Cf-4* and *Cf-9* resistance genes against *Cladosporium fulvum* are conserved in wild tomato species. *Mol. Plant. Microbe Interact.*, **18**, 1011–1021.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K. and Battistuzzi, F.U. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.*, **35**, 1547–1549.

Lacombe, S., Rougon-Cardoso, A., Sherwood, E., et al. (2010) Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. *Nat. Biotechnol.*, **28**, 365–369.

Lange, O. de, Schreiber, T., Schandry, N., Radeck, J., Braun, K.H., Koszinowski, J., Heuer, H., Strauß, A. and Lahaye, T. (2013) Breaking the DNA-binding code of Ralstonia solanacearum TAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease. *New Phytol.*, **199**, 773–786.

Langford, A.N. (1948) Autogenous necrosis in tomatoes immune from *Cladosporium fulvum* Cooke. *Can. J. Res.*, **26c**, 35–64.

Lazar, C. (2015) *imputeLCMD: a collection of methods for left-censored missing data imputation*.

Lee, B., Park, Y.-S., Lee, S., Song, G.C. and Ryu, C.-M. (2016) Bacterial RNAs activate innate immunity in Arabidopsis. *New Phytol.*, **209**, 785–797.

Lee, D., Bourdais, G., Yu, G., Robatzek, S. and Coaker, G. (2015) Phosphorylation of the plant immune regulator RPM1-INTERACTING PROTEIN4 enhances plant plasma membrane H^+-ATPase activity and inhibits flagellin-triggered immune responses in Arabidopsis. *Plant Cell Online*, **27**, 2042–2056.

Leitch, I.J., Hanson, L., Lim, K.Y., Kovarik, A., Chase, M.W., Clarkson, J.J. and Leitch, A.R. (2008) The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Ann. Bot.*, **101**, 805–814.

Le Roux, C., Huet, G., Jauneau, A., et al. (2015) A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell*, **161**, 1074–1088.

Lewis, J.D., Lee, A.H.-Y., Hassan, J.A., et al. (2013) The Arabidopsis ZED1 pseudokinase is required for ZAR1-mediated immunity induced by the *Pseudomonas syringae* type III effector HopZ1a. *Proc. Natl. Acad. Sci.*, **110**, 18722–18727.

Lewis, J.D., Wu, R., Guttman, D.S. and Desveaux, D. (2010) Allele-specific virulence attenuation of the *Pseudomonas syringae* HopZ1a type III effector via the Arabidopsis ZAR1 resistance protein. *PLoS Genet*, **6**, e1000894.

Li, M., Ma, X., Chiang, Y.-H., et al. (2014) Proline isomerization of the immune receptor-interacting protein RIN4 by a cyclophilin inhibits effector-triggered immunity in Arabidopsis. *Cell Host Microbe*, **16**, 473–483.

Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y.M., Buso, N. and Lopez, R. (2015) The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.*, **43**, W580–W584.

Lim, M.T.S. and Kunkel, B.N. (2004) The *Pseudomonas syringae* type III effector AvrRpt2 promotes virulence independently of RIN4, a predicted virulence target in *Arabidopsis thaliana*. *Plant J.*, **40**, 790–798.

Lin, T., Zhu, G., Zhang, J., et al. (2014) Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.*, **46**, 1220–1226.

Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K. and Chen, L.-L. (2017) CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. *Mol. Plant*, **10**, 530–532.

Liu, J., Elmore, J.M., Fuglsang, A.T., Palmgren, M.G., Staskawicz, B.J. and Coaker, G. (2009) RIN4 functions with plasma membrane H^+-ATPases to regulate stomatal apertures during pathogen attack. *PLOS Biol.*, **7**, e1000139.

Liu, J., Elmore, J.M., Lin, Z.-J.D. and Coaker, G. (2011) A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. *Cell Host Microbe*, **9**, 137–146.

Liu, Q., Yuan, M., Zhou, Y., Li, X., Xiao, J. and Wang, S. (2011) A paralog of the MtN3/saliva family recessively confers race-specific resistance to *Xanthomonas oryzae* in rice. *Plant Cell Environ.*, **34**, 1958–1969.

Liu, X., Grabherr, H.M., Willmann, R., et al. (2014) Host-induced bacterial cell wall decomposition mediates pattern-triggered immunity in Arabidopsis. *eLife*, **3**.

Liu, Y., Schiff, M., Marathe, R. and Dinesh-Kumar, S.P. (2002) Tobacco *Rar1*, *EDS1* and *NPR1*/*NIM1* like genes are required for *N*-mediated resistance to tobacco mosaic virus. *Plant J. Cell Mol. Biol.*, **30**, 415–429.

Lorenzo, G.D., Cervone, F., Bellincampi, D., et al. (1994) Polygalacturonase, PGIP and oligogalacturonides in cell-cell communication. *Biochem. Soc. Trans.*, **22**, 394–397.

Lozano-Torres, J.L., Wilbers, R.H.P., Gawronski, P., 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. *Proc. Natl. Acad. Sci.*, **109**, 10119–10124.

Lozano-Torres, J.L., Wilbers, R.H.P., Warmerdam, S., et al. (2014) Apoplastic venom allergen-like proteins of cyst nematodes modulate the activation of basal plant innate immunity by cell surface receptors. *PLoS Pathog*, **10**, e1004569.

Lu, R., Malcuit, I., Moffett, P., et al. (2003) High throughput virus-induced gene silencing implicates heat shock protein 90 in plant disease resistance. *EMBO J.*, **22**, 5690–5699.

Lu, X., Kracher, B., Saur, I.M.L., Bauer, S., Ellwood, S.R., Wise, R., Yaeno, T., Maekawa, T. and Schulze-Lefert, P. (2016) Allelic barley MLA immune receptors recognize sequence-unrelated avirulence effectors of the powdery mildew pathogen. *Proc. Natl. Acad. Sci.*, **113**, E6486–E6495.

Lu, Y., Hatsugai, N., Katagiri, F., Ishimaru, C.A. and Glazebrook, J. (2015) Putative serine protease effectors of *Clavibacter michiganensis* induce a hypersensitive response in the apoplast of *Nicotiana* species. *Mol. Plant. Microbe Interact.*, **28**, 1216–1226.

Luderer, R., Rivas, S., Nürnberger, T., et al. (2001) No evidence for binding between resistance gene product Cf-9 of tomato and avirulence gene product AVR9 of *Cladosporium fulvum*. *Mol. Plant. Microbe Interact.*, **14**, 867–876.

Luderer, R., Takken, F.L.W., Wit, P.J.G.M. de and Joosten, M.H.A.J. (2002) *Cladosporium fulvum* overcomes *Cf-2*-mediated resistance by producing truncated AVR2 elicitor proteins. *Mol. Microbiol.*, **45**, 875–884.

Ma, Z., Song, T., Zhu, L., et al. (2015) A *Phytophthora sojae* glycoside hydrolase 12 protein is a major virulence factor during soybean infection and Is recognized as a PAMP. *Plant Cell*, **27**, 2057–2072.

Ma, Z., Zhu, L., Song, T., et al. (2017) A paralogous decoy protects *Phytophthora sojae* apoplastic effector *Ps*XEG1 from a host inhibitor. *Science*, **355**, 710–714.

Mackey, D., Holt III, B.F., Wiig, A. and Dangl, J.L. (2002) RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for *RPM1*-Mediated resistance in Arabidopsis. *Cell*, **108**, 743–754.

Maqbool, A., Saitoh, H., Franceschetti, M., Stevenson, C.E.M., Uemura, A., Kanzaki, H., Kamoun, S., Terauchi, R. and Banfield, M.J. (2015) Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife*, **4**, e08709.

Martin, G.B., Brommonschenkel, S.H., Chunwongse, J., Frary, A., Ganal, M.W., Spivey, R., Wu, T., Earle, E.D. and Tanksley, S.D. (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science*, **262**, 1432–1436.

Martin, G.B., Frary, A., Wu, T., Brommonschenkel, S., Chunwongse, J., Earle, E.D. and Tanksley, S.D. (1994) A member of the tomato *Pto* gene family confers sensitivity to fenthion resulting in rapid cell death. *Plant Cell*, **6**, 1543–1552.

Mesarich, C.H., Griffiths, S.A., van der Burgt, A., Ökmen, B., Beenen, H.G., Etalo, D.W., Joosten, M.H.A.J. and de Wit, P.J.G.M. (2014) Transcriptome sequencing uncovers the *Avr5* avirulence gene of the tomato leaf mold pathogen *Cladosporium fulvum*. *Mol. Plant. Microbe Interact.*, **27**, 846–857.

Mesarich, C.H., Stergiopoulos, I., Beenen, H.G., Cordovez, V., Guo, Y., Karimi Jashni, M., Bradshaw, R.E. and de Wit, P.J.G.M. (2015) A conserved proline residue in Dothideomycete Avr4 effector proteins is required to trigger a Cf-4-dependent hypersensitive response. *Mol. Plant Pathol.*, n/a-n/a.

Michalski, A., Damoc, E., Hauschild, J.-P., et al. (2011) Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics*, **10**, M111.011015.

Mindrinos, M., Katagiri, F., Yu, G.-L. and Ausubel, F.M. (1994) The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell*, **78**, 1089–1099.

Misas-Villamil, J.C. and Hoorn, R.A.L. van der (2008) Enzyme–inhibitor interactions at the plant–pathogen interface. *Curr. Opin. Plant Biol.*, **11**, 380–388.

Moore, J.W., Herrera-Foessel, S., Lan, C., et al. (2015) A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. *Nat. Genet.*, **47**, 1494–1498.

Mueller, A.N., Ziemann, S., Treitschke, S., Aßmann, D. and Doehlemann, G. (2013) Compatibility in the *Ustilago maydis*–maize interaction requires inhibition of host cysteine proteases by the fungal effector Pit2. *PLOS Pathog.*, **9**, e1003177.

Mukhtar, M.S., Carvunis, A.-R., Dreze, M., et al. (2011) Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science*, **333**, 596–601.

Mysore, K.S. and Ryu, C.-M. (2004) Nonhost resistance: how much do we know? *Trends Plant Sci.*, **9**, 97–104.

Naim, F., Nakasugi, K., Crowhurst, R.N., Hilario, E., Zwart, A.B., Hellens, R.P., Taylor, J.M., Waterhouse, P.M. and Wood, C.C. (2012) Advanced engineering of lipid metabolism in *Nicotiana benthamiana* using a draft genome and the V2 viral silencing-suppressor protein. *PLOS ONE*, **7**, e52717.

Nakasugi, K., Crowhurst, R., Bally, J. and Waterhouse, P. (2014) Combining transcriptome assemblies from multiple *de novo* assemblers in the allo-tetraploid plant *Nicotiana benthamiana*. *PLOS ONE*, **9**, e91776.

Nekrasov, V., Staskawicz, B.J., Weigel, D., Jones, J.D.G. and Kamoun, S. (2013) Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.*, **31**, 691.

Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D. and Kamoun, S. (2017) Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci. Rep.*, **7**, 482.

Nirmala, J., Drader, T., Lawrence, P.K., et al. (2011) Concerted action of two avirulent spore effectors activates *Reaction to Puccinia graminis 1* (*Rpg1*)-mediated cereal stem rust resistance. *Proc. Natl. Acad. Sci.*, **108**, 14676–14681.

Ntoukakis, V., Saur, I.M., Conlan, B. and Rathjen, J.P. (2014) The changing of the guard: the Pto/Prf receptor complex of tomato and pathogen recognition. *Curr. Opin. Plant Biol.*, **20**, 69–74.

Offord, V., Coffey, T.J. and Werling, D. (2010) LRRfinder: a web application for the identification of leucine-rich repeats and an integrative Toll-like receptor database. *Dev. Comp. Immunol.*, **34**, 1035–1041.

Ohm, R.A., Feau, N., Henrissat, B., et al. (2012) Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes* fungi. *PLOS Pathog.*, **8**, e1003037.

Olsen, J.V., Godoy, L.M.F. de, Li, G., et al. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics*, **4**, 2010–2021.

Oome, S., Raaymakers, T.M., Cabral, A., Samwel, S., Böhm, H., Albert, I., Nürnberger, T. and Van den Ackerveken, G.F.J.M. (2014) Nep1-like proteins from three kingdoms of life act as a microbe-associated molecular pattern in Arabidopsis. *Proc. Natl. Acad. Sci.*, **111**, 16955–16960.

Peart, J.R., Mestre, P., Lu, R., Malcuit, I. and Baulcombe, D.C. (2005) NRG1, a CC-NB-LRR Protein, together with N, a TIR-NB-LRR Protein, mediates resistance against tobacco mosaic virus. *Curr. Biol.*, **15**, 968–973.

Peele, H.M., Guan, N., Fogelqvist, J. and Dixelius, C. (2014) Loss and retention of resistance genes in five species of the Brassicaceae family. *BMC Plant Biol.*, **14**, 298.

Petersen, T.N., Brunak, S., Heijne, G. von and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, **8**, 785–786.

Phipson, B., Lee, S., Majewski, I.J., Alexander, W.S. and Smyth, G.K. (2016) Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.*, **10**, 946–963.

Piffanelli, P., Zhou, F., Casais, C., Orme, J., Jarosch, B., Schaffrath, U., Collins, N.C., Panstruga, R. and Schulze-Lefert, P. (2002) The barley MLO modulator of defense and cell death Is responsive to biotic and abiotic stress stimuli. *Plant Physiol.*, **129**, 1076–1085.

Pink, M., Verma, N., Rettenmeier, A.W. and Schmitz-Spanke, S. (2010) CBB staining protocol with higher sensitivity and mass spectrometric compatibility. *ELECTROPHORESIS*, **31**, 593–598.

Pruitt, R.N., Joe, A., Zhang, W., Feng, W., Stewart, V., Schwessinger, B., Dinneny, J.R. and Ronald, P.C. (2017) A microbially derived tyrosine-sulfated peptide mimics a plant peptide hormone. *New Phytol.*, **215**, 725–736.

Qi, D., DeYoung, B.J. and Innes, R.W. (2012) Structure-function analysis of the coiled-coil and leucine-rich repeat domains of the RPS5 disease resistance protein. *Plant Physiol.*, **158**, 1819–1832.

Qi, D., Dubiella, U., Kim, S.H., Isaiah, S., Dowen, R.H., Dixon, J.E. and Innes, R.W. (2014) Recognition of the protein kinase AVRPPHB SUSCEPTIBLE1 by the disease resistance protein RESISTANCE TO PSEUDOMONAS SYRINGAE5 Is dependent on S-acylation and an exposed loop in AVRPPHB SUSCEPTIBLE. *Plant Physiol.*, **164**, 340–351.

Qi, Dong, Dubiella, U., Kim, S.H., Sloss, D.I., Dowen, R.H., Dixon, J.E. and Innes, R.W. (2014) Recognition of the protein kinase AVRPPHB SUSCEPTIBLE1 by the disease resistance protein RESISTANCE TO PSEUDOMONAS SYRINGAE5 is dependent on S-acylation and an exposed loop in AVRPPHB SUSCEPTIBLE1. *Plant Physiol.*, **164**, 340–351.

Qin, C., Yu, C., Shen, Y., et al. (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 5135–5140.

Rappsilber, J., Mann, M. and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.*, **2**, 1896–1906.

Ravensdale, M., Bernoux, M., Ve, T., Kobe, B., Thrall, P.H., Ellis, J.G. and Dodds, P.N. (2012) Intramolecular interaction influences binding of the flax L5 and L6 resistance proteins to their AvrL567 Ligands. *PLoS Pathog*, **8**, e1003004.

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A. and Finn, R.D. (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.*, **46**, D624–D632.

Read, A.C., Rinaldi, F.C., Hutin, M., He, Y.-Q., Triplett, L.R. and Bogdanove, A.J. (2016) Suppression of *Xo1*-mediated disease resistance in rice by a truncated, non-DNA-binding TAL effector of *Xanthomonas oryzae*. *Front. Plant Sci.*, **7**, 1516.

Richau, K.H., Kaschani, F., Verdoes, M., et al. (2012) Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics. *Plant Physiol.*, **158**, 1583–1599.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47–e47.

Robatzek, S., Bittel, P., Chinchilla, D., Köchner, P., Felix, G., Shiu, S.-H. and Boller, T. (2007) Molecular identification and characterization of the tomato flagellin receptor LeFLS2, an orthologue of Arabidopsis FLS2 exhibiting characteristically different perception specificities. *Plant Mol. Biol.*, **64**, 539–547.

Römer, P., Hahn, S., Jordan, T., Strauß, T., Bonas, U. and Lahaye, T. (2007) Plant pathogen recognition mediated by promoter activation of the pepper *Bs3* resistance gene. *Science*, **318**, 645–648.

Römer, P., Recht, S. and Lahaye, T. (2009) A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. *Proc. Natl. Acad. Sci.*, **106**, 20526–20531.

Rommens, C.M., Salmeron, J.M., Oldroyd, G.E. and Staskawicz, B.J. (1995) Intergeneric transfer and functional expression of the tomato disease resistance gene *Pto*. *Plant Cell*, **7**, 1537–1544.

Ron, M. and Avni, A. (2004) The receptor for the fungal elicitor ethylene-inducing xylanase is a member of a resistance-like gene family in tomato. *Plant Cell*, **16**, 1604–1615.

Rooney, H.C.E., van't Klooster, J.W., van der Hoorn, R.A.L., Joosten, M.H.A.J., Jones, J.D.G. and de Wit, P.J.G.M. (2005) *Cladosporium* Avr2 inhibits tomato Rcr3 protease required for *Cf-2*-dependent disease resistance. *Science*, **308**, 1783–1786.

Rosebrock, T.R., Zeng, L., Brady, J.J., Abramovitch, R.B., Xiao, F. and Martin, G.B. (2007) A bacterial E3 ubiquitin ligase targets a host protein kinase to disrupt plant immunity. *Nature*, **448**, 370–374.

Rotblat, B., Enshell-Seijffers, D., Gershoni, J.M., Schuster, S. and Avni, A. (2002) Identification of an essential component of the elicitation active site of the EIX protein elicitor. *Plant J. Cell Mol. Biol.*, **32**, 1049–1055.

Russell, A.R., Ashfield, T. and Innes, R.W. (2015) *Pseudomonas syringae* effector AvrPphB suppresses AvrB-induced activation of RPM1 but not AvrRpm1-induced activation. *Mol. Plant. Microbe Interact.*, **28**, 727–735.

Salmeron, J.M., Oldroyd, G.E.D., Rommens, C.M.T., Scofield, S.R., Kim, H.-S., Lavelle, D.T., Dahlbeck, D. and Staskawicz, B.J. (1996) Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. *Cell*, **86**, 123–133.

Sarris, P.F., Cevik, V., Dagdas, G., Jones, J.D.G. and Krasileva, K.V. (2016) Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biol.*, **14**, 8.

Sarris, P.F., Duxbury, Z., Huh, S.U., et al. (2015) A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell*, **161**, 1089–1100.

Saucet, S.B., Ma, Y., Sarris, P.F., Furzer, O.J., Sohn, K.H. and Jones, J.D.G. (2015) Two linked pairs of Arabidopsis TNL resistance genes independently confer recognition of bacterial effector AvrRps4. *Nat. Commun.*, **6**, 6338.

Schardon, K., Hohl, M., Graff, L., Pfannstiel, J., Schulze, W., Stintzi, A. and Schaller, A. (2016) Precursor processing for plant peptide hormone maturation by subtilisin-like serine proteinases. *Science*, aai8550.

van Schie, C.C.N. and Takken, F.L.W. (2014) Susceptibility genes 101: how to be a good host. *Annu. Rev. Phytopathol.*, **52**, 551–581.

Schiermeyer, A., Hartenstein, H., Mandal, M.K., Otte, B., Wahner, V. and Schillberg, S. (2009) A membrane-bound matrix-metalloproteinase from *Nicotiana tabacum* cv. BY-2 is induced by bacterial pathogens. *BMC Plant Biol.*, **9**, 83.

Schilling, O. and Overall, C.M. (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat. Biotechnol.*, **26**, 685–694.

Schornack, S., Ballvora, A., Gürlebeck, D., Peart, J., Baulcombe, D., Ganal, M., Baker, B., Bonas, U. and Lahaye, T. (2004) The tomato resistance protein Bs4 is a predicted non-nuclear TIR-NB-LRR protein

that mediates defense responses to severely truncated derivatives of AvrBs4 and overexpressed AvrBs3. *Plant J. Cell Mol. Biol.*, **37**, 46–60.

Schultink, A., Qi, T., Bally, J. and Staskawicz, B.J. (2018) Using forward genetics in *Nicotiana benthamiana* to uncover the immune signaling pathway mediating recognition of the *Xanthomonas perforans* effector XopJ4. *New Phytol.*, **0**.

Schulze-Lefert, P. and Panstruga, R. (2011) A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends Plant Sci.*, **16**, 117–125.

Seear, P.J. and Dixon, M.S. (2003) Variable leucine-rich repeats of tomato disease resistance genes *Cf-2* and *Cf-5* determine specificity. *Mol. Plant Pathol.*, **4**, 199–202.

Selote, D. and Kachroo, A. (2010) *RPG1-B*-derived resistance to AvrB-expressing *Pseudomonas syringae* requires RIN4-like proteins in soybean. *Plant Physiol.*, **153**, 1199–1211.

Senthil-Kumar, M. and Mysore, K.S. (2014) Tobacco rattle virus–based virus-induced gene silencing in *Nicotiana benthamiana*. *Nat. Protoc.*, **9**, 1549–1562.

Senthil-Kumar, M., Wang, M., Chang, J., et al. (2018) Virus-induced gene silencing database for phenomics and functional genomics in *Nicotiana benthamiana*. *Plant Direct*, **2**, e00055.

Seto, D., Koulena, N., Lo, T., Menna, A., Guttman, D.S. and Desveaux, D. (2017) Expanded type III effector recognition by the ZAR1 NLR protein using ZED1-related kinases. *Nat. Plants*, **3**, 17027.

Shabab, M., Shindo, T., Gu, C., et al. (2008) Fungal effector protein AVR2 targets diversifying defense-related cys proteases of tomato. *Plant Cell Online*, **20**, 1169–1183.

Shao, F., Golstein, C., Ade, J., Stoutemyer, M., Dixon, J.E. and Innes, R.W. (2003) Cleavage of Arabidopsis PBS1 by a bacterial type III effector. *Science*, **301**, 1230–1233.

Shen, Q.-H., Zhou, F., Bieri, S., Haizel, T., Shirasu, K. and Schulze-Lefert, P. (2003) Recognition specificity and RAR1/SGT1 dependence in barley *Mla* disease resistance genes to the powdery mildew fungus. *Plant Cell*, **15**, 732–744.

Shindo, T., Kaschani, F., Yang, F., et al. (2016) Screen of non-annotated small secreted proteins of *Pseudomonas syringae* reveals a virulence factor that inhibits tomato immune proteases. *PLOS Pathog*, **12**, e1005874.

Sierro, N., Battey, J.N., Ouadi, S., Bovet, L., Goepfert, S., Bakaher, N., Peitsch, M.C. and Ivanov, N.V. (2013) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol.*, **14**, R60.

Sierro, N., Battey, J.N.D., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., Goepfert, S., Peitsch, M.C. and Ivanov, N.V. (2014) The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.*, **5**, 3833.

Sievers, F., Wilm, A., Dineen, D., et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212.

Sindhu, A., Chintamanani, S., Brandt, A.S., Zanis, M., Scofield, S.R. and Johal, G.S. (2008) A guardian of grasses: specific origin and conservation of a unique disease-resistance gene in the grass lineage. *Proc. Natl. Acad. Sci.*, **105**, 1762–1767.

Smakowska-Luzan, E., Mott, G.A., Parys, K., et al. (2018) An extracellular network of *Arabidopsis* leucine-rich repeat receptor kinases. *Nature*, **553**, 342.

Song, J., Win, J., Tian, M., Schornack, S., Kaschani, F., Ilyas, M., Hoorn, R.A.L. van der and Kamoun, S. (2009) Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3. *Proc. Natl. Acad. Sci.*, **106**, 1654–1659.

Sparkes, I.A., Runions, J., Kearns, A. and Hawes, C. (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat. Protoc.*, **1**, 2019–2025.

Spassieva, S.D., Markham, J.E. and Hille, J. (2002) The plant disease resistance gene *Asc-1* prevents disruption of sphingolipid metabolism during AAL-toxin-induced programmed cell death. *Plant J.*, **32**, 561–572.

Sperschneider, J., Dodds, P.N., Singh, K.B. and Taylor, J.M. (2018) ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. *New Phytol.*, **217**, 1764–1778.

Srivastava, R., Liu, J.-X., Guo, H., Yin, Y. and Howell, S.H. (2009) Regulation and processing of a plant peptide hormone, AtRALF23, in Arabidopsis. *Plant J.*, **59**, 930–939.

Srivastava, R., Liu, J.-X. and Howell, S.H. (2008) Proteolytic processing of a precursor protein for a growth-promoting peptide by a subtilisin serine protease in Arabidopsis. *Plant J.*, **56**, 219–227.

Stanke, M., Schöffmann, O., Morgenstern, B. and Waack, S. (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*, **7**, 62.

Steinbrenner, A.D., Goritschnig, S. and Staskawicz, B.J. (2015) Recognition and activation domains contribute to allele-specific responses of an Arabidopsis NLR receptor to an oomycete effector protein. *PLOS Pathog.*, **11**, e1004665.

Stemmer, M., Thumberger, T., Keyer, M. del S., Wittbrodt, J. and Mateo, J.L. (2015) CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLOS ONE*, **10**, e0124633.

Stergiopoulos, I., van den Burg, H.A., Ökmen, B., Beenen, H.G., van Liere, S., Kema, G.H.J. and de Wit, P.J.G.M. (2010) Tomato Cf resistance proteins mediate recognition of cognate homologous effectors from fungi pathogenic on dicots and monocots. *Proc. Natl. Acad. Sci.*, **107**, 7610–7615.

Stergiopoulos, I., Kock, M.J.D. de, Lindhout, P. and de Wit, P.J.G.M. (2007) Allelic variation in the effector genes of the tomato pathogen *Cladosporium fulvum* reveals different modes of adaptive evolution. *Mol. Plant. Microbe Interact.*, **20**, 1271–1283.

Steuernagel, B., Periyannan, S.K., Hernández-Pinzón, I., et al. (2016) Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.*, **34**, 652–655.

Stoger, E., Fischer, R., Moloney, M. and Ma, J.K.-C. (2014) Plant molecular pharming for the treatment of chronic and infectious diseases. *Annu. Rev. Plant Biol.*, **65**, 743–768.

Strauß, T., Poecke, R.M.P. van, Strauß, Annett, et al. (2012) RNA-seq pinpoints a *Xanthomonas* TAL-effector activated resistance gene in a large-crop genome. *Proc. Natl. Acad. Sci.*, **109**, 19480–19485.

Sueldo, D., Ahmed, A., Misas-Villamil, J., Colby, T., Tameling, W., Joosten, M.H.A.J. and van der Hoorn, R.A.L. (2014) Dynamic hydrolase activities precede hypersensitive tissue collapse in tomato seedlings. *New Phytol.*, **203**, 913–925.

Sun, Y., Li, L., Macho, A.P., Han, Z., Hu, Z., Zipfel, C., Zhou, J.-M. and Chai, J. (2013) Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. *Science*, **342**, 624–628.

Swiderski, M.R. and Innes, R.W. (2001) The Arabidopsis *PBS1* resistance gene encodes a member of a novel protein kinase subfamily. *Plant J.*, **26**, 101–112.

Takken, F.L.W., Thomas, C.M., Joosten, M.H.A.J., Golstein, C., Westerink, N., Hille, J., Nijkamp, H.J.J., Wit, P.J.G.M. de and Jones, J.D.G. (1999) A second gene at the tomato *Cf-4* locus confers resistance to *Cladosporium fulvum* through recognition of a novel avirulence determinant. *Plant J.*, **20**, 279–288.

Tang, J., Han, Z., Sun, Y., Zhang, H., Gong, X. and Chai, J. (2015) Structural basis for recognition of an endogenous peptide by the plant receptor kinase PEPR1. *Cell Res.*, **25**, 110–120.

Taylor, A. and Qiu, Y.-L. (2017) Evolutionary history of subtilases in land plants and their involvement in symbiotic interactions. *Mol. Plant. Microbe Interact.*, **30**, 489–501.

Thomas, C.M., Jones, D.A., Parniske, M., Harrison, K., Balint-Kurti, P.J., Hatzixanthis, K. and Jones, J.D.G. (1997) Characterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in Cf-4 and Cf-9. *Plant Cell*, **9**, 2209–2224.

Thomma, B.P.H.J., Esse, H.P.V., Crous, P.W. and Wit, P.J.G.M. de (2005) *Cladosporium fulvum* (syn. *Passalora fulva*), a highly specialized plant pathogen as a model for functional studies on plant pathogenic Mycosphaerellaceae. *Mol. Plant Pathol.*, **6**, 379–393.

Thomma, B.P.H.J., Nürnberger, T. and Joosten, M.H.A.J. (2011) Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *Plant Cell*, **23**, 4–15.

Tian, D., Wang, J., Zeng, X., et al. (2014) The rice TAL effector–dependent resistance protein XA10 triggers cell death and calcium depletion in the endoplasmic reticulum. *Plant Cell*, **26**, 497–515.

Tian, M., Benedetti, B. and Kamoun, S. (2005) A second Kazal-like protease inhibitor from *Phytophthora infestans* inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato. *Plant Physiol.*, **138**, 1785–1793.

Tian, M., Huitema, E., Cunha, L. da, Torto-Alalibo, T. and Kamoun, S. (2004) A Kazal-like extracellular serine protease inhibitor from *Phytophthora infestans* targets the tomato pathogenesis-related protease P69B. *J. Biol. Chem.*, **279**, 26370–26377.

Tian, M., Win, J., Song, J., Hoorn, R. van der, van der Knaap, E. and Kamoun, S. (2007) A Phytophthora infestans Cystatin-Like Protein Targets a Novel Tomato Papain-Like Apoplastic Protease. *Plant Physiol.*, **143**, 364–377.

Tornero, P., Mayda, E., Gómez, M.D., Cañas, L., Conejero, V. and Vera, P. (1996) Characterization of LRP, a leucine-rich repeat (LRR) protein from tomato plants that is processed during pathogenesis. *Plant J.*, **10**, 315–330.

Trdá, L., Fernandez, O., Boutrot, F., et al. (2014) The grapevine flagellin receptor VvFLS2 differentially recognizes flagellin-derived epitopes from the endophytic growth-promoting bacterium *Burkholderia phytofirmans* and plant pathogenic bacteria. *New Phytol.*, **201**, 1371–1384.

Triplett, L.R., Cohen, S.P., Heffelfinger, C., Schmidt, C.L., Huerta, A.I., Tekete, C., Verdier, V., Bogdanove, A.J. and Leach, J.E. (2016) A resistance locus in the American heirloom rice variety Carolina Gold Select is triggered by TAL effectors with diverse predicted targets and is effective against African strains of *Xanthomonas oryzae* pv. *oryzicola*. *Plant J. Cell Mol. Biol.*, **87**, 472–483.

Truniger, V. and Aranda, M.A. (2009) Recessive resistance to plant viruses. In G. L. and J. P. Carr, ed. *Advances in Virus Research*. Natural and Engineered Resistance to Plant Viruses, Part I. Academic Press, pp. 119–231.

Tyanova, S., Temu, T. and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.*, **11**, 2301–2319.

Verlaan, M.G., Hutton, S.F., Ibrahem, R.M., Kormelink, R., Visser, R.G.F., Scott, J.W., Edwards, J.D. and Bai, Y. (2013) The Tomato Yellow Leaf Curl Virus resistance genes *Ty-1* and *Ty-3* are allelic and code for DFDGD-class RNA–dependent RNA polymerases. *PLOS Genet.*, **9**, e1003399.

Vossen, J.H., Arkel, G. van, Bergervoet, M., Jo, K.-R., Jacobsen, E. and Visser, R.G.F. (2016) The *Solanum demissum R8* late blight resistance gene is an *Sw-5* homologue that has been deployed worldwide in late blight resistant varieties. *Theor. Appl. Genet.*, **129**, 1785–1796.

Wang, C., Zhang, X., Fan, Y., et al. (2015) XA23 Is an executor R protein and confers broad-spectrum disease resistance in rice. *Mol. Plant*, **8**, 290–302.

Wang, C.-I.A., Gunčar, G., Forwood, J.K., et al. (2007) Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *Plant Cell*, **19**, 2898–2912.

Wang, G., Roux, B., Feng, F., et al. (2015) The decoy substrate of a pathogen effector and a pseudokinase specify pathogen-induced modified-self recognition and immunity in plants. *Cell Host Microbe*, **18**, 285–295.

Wang, X. and Bennetzen, J.L. (2015) Current status and prospects for the study of *Nicotiana* genomics, genetics, and nicotine biosynthesis genes. *Mol. Genet. Genomics*, **290**, 11–21.

Warren, R.F., Henk, A., Mowery, P., Holub, E. and Innes, R.W. (1998) A mutation within the leucine-rich repeat domain of the Arabidopsis disease resistance gene *RPS5* partially suppresses multiple bacterial and downy mildew resistance genes. *Plant Cell*, **10**, 1439–1452.

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V. and Zdobnov, E.M. (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.*, **35**, 543–548.

Weber, E., Engler, C., Gruetzner, R., Werner, S. and Marillonnet, S. (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE*, **6**, e16765.

Wei, C.-F., Kvitko, B.H., Shimizu, R., Crabill, E., Alfano, J.R., Lin, N.-C., Martin, G.B., Huang, H.-C. and Collmer, A. (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*. *Plant J.*, **51**, 32–46.

Westerink, N., Brandwagt, B.F., Wit, P.J.G.M. de and Joosten, M.H.A.J. (2004) *Cladosporium fulvum* circumvents the second functional resistance gene homologue at the *Cf-4* locus (*Hcr9-4E*) by secretion of a stable avr4E isoform. *Mol. Microbiol.*, **54**, 533–545.

Whelan, S. and Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.*, **18**, 691–699.

Whitham, S., Dinesh-Kumar, S.P., Choi, D., Hehl, R., Corr, C. and Baker, B. (1994) The product of the tobacco mosaic virus resistance gene *N*: similarity to toll and the interleukin-1 receptor. *Cell*, **78**, 1101–1115.

Wilf, P., Carvalho, M.R., Gandolfo, M.A. and Cúneo, N.R. (2017) Eocene lantern fruits from Gondwanan Patagonia and the early origins of Solanaceae. *Science*, **355**, 71–75.

Wilton, M., Subramaniam, R., Elmore, J., Felsensteiner, C., Coaker, G. and Desveaux, D. (2010) The type III effector $HopF2_{Pto}$ targets Arabidopsis RIN4 protein to promote *Pseudomonas syringae* virulence. *Proc. Natl. Acad. Sci.*, **107**, 2349–2354.

de Wit, P.J.G.M., Burgt, A. van der, Ökmen, B., et al. (2012) The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. *PLOS Genet.*, **8**, e1003088.

Witek, K., Jupe, F., Witek, A.I., Baker, D., Clark, M.D. and Jones, J.D.G. (2016) Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. *Nat. Biotechnol.*, **34**, 656–660.

Wrzaczek, M., Vainonen, J.P., Stael, S., et al. (2015) GRIM REAPER peptide binds to receptor kinase PRK5 to trigger cell death in Arabidopsis. *EMBO J.*, **34**, 55–66.

Wu, C.-H., Abd-El-Haliem, A., Bozkurt, T.O., Belhaj, K., Terauchi, R., Vossen, J.H. and Kamoun, S. (2017) NLR network mediates immunity to diverse plant pathogens. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, 8113–8118.

Wu, J., Kou, Y., Bao, J., et al. (2015) Comparative genomics identifies the *Magnaporthe oryzae* avirulence effector AvrPi9 that triggers Pi9-mediated blast resistance in rice. *New Phytol.*, **206**, 1463–1475.

Wulff, B.B.H., Thomas, C.M., Smoker, M., Grant, M. and Jones, J.D.G. (2001) Domain swapping and gene shuffling identify sequences required for induction of an Avr-dependent hypersensitive response by the tomato Cf-4 and Cf-9 proteins. *Plant Cell*, **13**, 255–272.

Xing, W., Zou, Y., Liu, Q., et al. (2007) The structural basis for activation of plant immunity by bacterial effector protein AvrPto. *Nature*, **449**, 243–247.

Xu, Q.-F., Cheng, W.-S., Li, S.-S., Li, W., Zhang, Z.-X., Xu, Y.-P., Zhou, X.-P. and Cai, X.-Z. (2012) Identification of genes required for *Cf*-dependent hypersensitive cell death by combined proteomic and RNA interfering analyses. *J. Exp. Bot.*, **63**, 2421–2435.

Xu, R.-Q., Blanvillain, S., Feng, J.-X., et al. (2008) $AvrAC_{Xcc8004}$, a type III effector with a leucine-rich repeat domain from *Xanthomonas campestris* pathovar *campestris* confers avirulence in vascular tissues of *Arabidopsis thaliana* ecotype Col-0. *J. Bacteriol.*, **190**, 343–355.

Xu, S., Brockmöller, T., Navarro-Quezada, A., et al. (2017) Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc. Natl. Acad. Sci.*, **114**, 6133–6138.

Yamada, K., Yamashita-Yamada, M., Hirase, T., Fujiwara, T., Tsuda, K., Hiruma, K. and Saijo, Y. (2016) Danger peptide receptor signaling in plants ensures basal immunity upon pathogen-induced depletion of BAK1. *EMBO J.*, **35**, 46–61.

Yuan, M., Ke, Y., Huang, R., Ma, L., Yang, Z., Chu, Z., Xiao, J., Li, X. and Wang, S. (2016) A host basal transcription factor is a key component for infection of rice by TALE-carrying bacteria. *eLife*, **5**, e19605.

Zeng, X., Tian, D., Gu, K., Zhou, Z., Yang, X., Luo, Y., White, F.F. and Yin, Z. (2015) Genetic engineering of the *Xa10* promoter for broad-spectrum and durable resistance to *Xanthomonas oryzae* pv. *oryzae*. *Plant Biotechnol. J.*, **13**, 993–1001.

Zhang, J., Li, W., Xiang, T., et al. (2010) Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a *Pseudomonas syringae* effector. *Cell Host Microbe*, **7**, 290–301.

Zhang, J., Yin, Z. and White, F. (2015) TAL effectors and the executor *R* genes. *Front. Plant Sci.*, **6**.

Zhang, L., Kars, I., Essenstam, B., et al. (2014) Fungal endopolygalacturonases are recognized as microbe-associated molecular patterns by the Arabidopsis receptor-like protein RESPONSIVENESS TO BOTRYTIS POLYGALACTURONASES1. *Plant Physiol.*, **164**, 352–364.

Zhang, Z., Liu, Y., Huang, H., Gao, M., Wu, D., Kong, Q. and Zhang, Y. (2017) The NLR protein SUMM2 senses the disruption of an immune signaling MAP kinase cascade via CRCK3. *EMBO Rep.*, **18**, 292–302.

Zhang, Z., Wu, Y., Gao, M., Zhang, J., Kong, Q., Liu, Y., Ba, H., Zhou, J. and Zhang, Y. (2012) Disruption of PAMP-induced MAP kinase cascade by a *Pseudomonas syringae* effector activates plant immunity mediated by the NB-LRR protein SUMM2. *Cell Host Microbe*, **11**, 253–263.

Zhao, T., Jiang, J., Liu, G., He, S., Zhang, H., Chen, X., Li, J. and Xu, X. (2016) Mapping and candidate gene screening of tomato *Cladosporium fulvum*-resistant gene *Cf-19*, based on high-throughput sequencing technology. *BMC Plant Biol.*, **16**, 51.

Zhu, M., Jiang, L., Bai, B., et al. (2017) The intracellular immune receptor Sw-5b confers broad-spectrum resistance to tospoviruses through recognition of a conserved 21-amino acid viral effector epitope. *Plant Cell*, **29**, 2214–2232.

Zhu, M., Shao, F., Innes, R.W., Dixon, J.E. and Xu, Z. (2004) The crystal structure of *Pseudomonas* avirulence protein AvrPphB: a papain-like fold with a distinct substrate-binding site. *Proc. Natl. Acad. Sci.*, **101**, 302–307.

Ziemann, S., Linde, K. van der, Lahrmann, U., et al. (2018) An apoplastic peptide activates salicylic acid signalling in maize. *Nat. Plants*, **4**, 172–180.

Zimmermann, D., Gomez-Barrera, J.A., Pasule, C., Brack-Frick, U.B., Sieferer, E., Nicholson, T.M., Pfannstiel, J., Stintzi, A. and Schaller, A. (2016) Cell death control by matrix metalloproteinases. *Plant Physiol.*, **171**, 1456–1469.

Zipfel, C. (2014) Plant pattern-recognition receptors. *Trends Immunol.*, **35**, 345–351.