Adaptive evolution of rice immune receptors

Aleksandra Białas

Thesis submitted to the University of East Anglia
for the Degree of Doctor of Philosophy

October 2020

Abstract

Plants have evolved intracellular immune receptors, known as nucleotide-binding domain leucine-rich repeat-containing (NLR) proteins, that trigger effective immune responses upon perception of pathogen-derived effectors. Although typical NLRs share a conserved multidomain architecture, some of them carry unconventional so-called integrated domains that appear to have evolved from host targets of pathogen effectors. One example of such NLR is Pik-1—a rice immune receptor that confers disease resistance to the blast fungus *Magnaporthe oryzae*. Pik-1 carries a heavy metal–associated (HMA) domain that directly binds the AVR-Pik effector from *M. oryzae* by mimicking the host targets of this effector. In this thesis, I aimed to understand the evolutionary history of the Pik-1 receptor and its partner, Pik-2, and test hypotheses about adaptive evolution of the Pik-1–integrated HMA domain. Phylogenetic analyses of Pik orthologues revealed that the HMA domain integrated into Pik-1 before Oryzinae speciation and has been subject to the strong diversifying selection. Ancestral HMA sequence reconstruction coupled with functional studies showed that different allelic variants of Pik-1, Pikp-1 and Pikm-1, convergently evolved to recognise AVR-PikD. Using biochemical and biophysical approaches, I functionally characterised two regions in the Pik-1 HMA that independently evolved towards high-affinity AVR-PikD binding from the weaker ancestral state. In both cases the HMA domain only recently acquired the capacity to bind the AVR-PikD effector with high affinity, indicating that for most of its evolutionary history the HMA had not been subject to selection pressure imposed by this blast effector. In addition, although Pikp-1 and Pikm-1 receptors evolved to produce similar phenotypic outcomes, they underwent different evolutionary trajectories to do so. These findings paint a complex picture of the mechanisms and evolutionary dynamics of NLR adaptation and provide a robust evolutionary framework that can contribute a more comprehensive understanding of plant–microbe systems.

Table of Contents

List of Figures

List of Tables

Abbreviations

ADR1	activated disease resistance 1
APAF-1	apoptotic protease-activating factor 1
ATR	*Arabidopsis thaliana*–recognized
AVR	avirulence protein
Bd	*Brachypodium distachyon*
BIK1	botrytis-induced kinase 1
BLAST	Basic Local Alignment Search Tool
BOP	Bambusoideae, Oryzoideae, Pooideae
CC	coiled-coil
CC_{EDVID}	EDVID-like coiled-coil
CC_{G10}	G10-like coiled-coil
CC_{R}	RPW8-like coiled-coil
CED-4	cell death protein 4
CNL	CC-NLR
Co-IP	co-immunoprecipitation
cryo-EM	cryogenic electron microscopy
Dg	*Dactylis glomerata*
d_N	nonsynonymous nucleotide substitution
dpi	days post infiltration
d_S	synonymous nucleotide substitution
EPIC1	extracellular protease inhibitor with cystatin-like domain 1
FLS2	flagellin sensing 2
GTR	generalised time-reversible
HD1	helical domain 1
HF	His-FLAG
HIPP	heavy metal–associated isoprenylated plant proteins
HMA	heavy metal–associated
HR	hypersensitive response
HR4	homologue of RPW8 4
HSP90	heat shock protein 90
ID	integrated domain
iToL	Interactive Tree of Life
JTT	Jones-Taylor-Thornton
LB	lysogeny broth
Lp	*Leersia perrieri*
LRR	leucine-rich repeat
MAMP	microbe-associated molecular pattern
MAX	*M. oryzae* Avrs and ToxB-like
MBP	maltose-binding protein
MHD	Met-His-Asp
ML	maximum likelihood
Mla	mildew resistance locus A
MRC1	major resistance cluster 1
MUSCLE	Multiple Sequence Comparison by Log-Expectation
MYA	million years ago

NB-ARC	nucleotide-binding adaptor shared by APAF-1, certain *R* gene products and CED-4
NBD	nucleotide-binding domain
NJ	neighbour-joining
NLR	nucleotide-binding domain leucine-rich repeat-containing
NLRC4	NLR family CARD domain–containing protein 4
NLS	nucleus localisation signal
NOD	nucleotide-binding and oligomerisation domain
NRC	NLR-required for cell death
NRG1	N requirement gene 1
NRIP	N receptor-interacting protein 1
Ob	*Oryza brachyantha*
Oglum	*Oryza glumaepatula*
Ol	*Oryza longistaminata*
Oniva	*Oryza nivara*
Opunc	*Oryza punctata*
Os	*Oryza sativa*
P/A	presence/absence
PACMAD	Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, Danthonioideae
PCR	polymerase chain reaction
PDB	Protein Data Bank
Pik	*Pyricularia oryzae* resistance
PM	powdery mildew resistance
PP	posterior probability
PPh	*P. syringae* pv. *phaseolicola*
R	resistance
RAR1	required for MLA12 resistance 1
RAxML	Randomized Axelerated Maximum Likelihood
RBA1	response to the bacterial type III effector protein HopBA1
RGA	resistance gene analogue
RGC16	resistance gene candidate
RGH2	resistance gene homologue
RIN4	RPM1-interacting protein 4
RLCK	receptor-like cytoplasmatic kinase
RLK	receptor-like kinase
RLP	receptor-like protein
Rmo1	resistance to *Magnaporthe oryzae* 1
Roq1	recognition of XopQ 1
Rp1	resistance to *Peronosclerospora sorghi* 1
RPM1	resistance to *Pseudomonas syringae* pv. *maculicola* 1
RPP	resistance to *Peronospora parasitica*
RPS2	resistance to *Pseudomonas syringae*
RRS1	resistance to *Ralstonia solanacearum* 1
RU	response unit
Rx	resistance to *Phytophthora infestans* and potato virus X
RXLR	Arg-X-Leu-Arg
Sb	*Sorghum bicolor*

SGT1	suppressor of G2 allele of SKP1
Si	*Setaria italica*
SPR	surface plasmon resonance
Sr	stem rust resistance
SSFR	super-structure forming repeat
STAND	signal-transduction ATPases with numerous domains
Ta	*Triticum aestivum*
TIR	Toll/interleukin receptor
TNL	TIR-NLR
WHD	winged helical domain
XopQ	*Xanthomonas* outer protein Q
Yr	yellow rust resistance
ZAR1	HopZ-activated resistance 1
zfBED	zinc-finger BED

Acknowledgements

First and foremost, I would like to thank Sophien Kamoun for his supervision over the course of my PhD, his patient guidance and encouragement. I am forever grateful and happily beholden to him for inspiring my interest in the evolution. I would also like to thank my secondary supervisor, Mark Banfield, for being an engaged, thoughtful, and encouraging member of my supervisory team.

This work wouldn't be possible without *many* people from The Sainsbury Laboratory and John Innes Centre who have contributed to this project in various ways and helped me along the way. I am indebted for all the insights, advice, assistance, and support I have been given over the years.

I feel incredibly lucky to have been able to work with each of the past and current members of the Kamoun Lab. Thank you for creating such an incredible, supportive, and stimulating environment, and for shaping my view of what makes a great scientist and a stellar colleague. Special thank you to the best officemate I could have wished for, Erin Zess—yes, I am confident that sum of all of our office chats add up to something productive. I am also grateful to Mauricio Contreras, who has taken over the role of my constant companion in the Kamoun Lab and did not disappoint. Finally, no words can express my appreciation to Lida Derevnina, who has been an inspiring role model.

I want to thank Erin, Lida, Mauricio, Thorsten Langner, and Adeline Harant who offered their time and provided careful and insightful edits on my thesis that helped to shape the final product.

To the past and current members of The Sainsbury Laboratory, whom I had privilege to interact with—thank you for providing an encouraging and collegial workplace. I would like to extent a special thank you to TSLytherins, for being such an incredible student community and making my time at TSL so much more fun and memorable.

Personally, I am extremely grateful for all the friendships I have made during my PhD and to the friendships that I brought with me. To my Writing Retreat companions, especially Pacia and Julia, thank you for making sure I don't work too hard during these final months.

His very own thank you to Łukasz, for all the support and making sure we never run out of coffee and Bread Source treats.

Lastly, I wouldn't be here without my fans—my family. Thank you to my extended family, especially my dear aunties for their sincere interest and enthusiasm. Special shout-out to Kitka for all her patience and support and to my brother for much-needed entertainment. Most of all, I want to thank my parents—all of this is thanks to you.

This PhD was funded by Norwich Research Park, Doctoral Training Partnership, BBSRC.

Chapter 1: General Introduction

1.1 Plant innate immunity—overview

During their lifetime, plants are constantly exposed to various biotic and abiotic stresses. Numerous pathogens—such as bacteria, viruses, fungi, and oomycetes, as well as parasites like nematodes, insects, and even parasitic plants—deploy elaborate strategies to successfully colonize their hosts and complete their life cycles. In order to defend themselves against intruders, plants have evolved an innate immune system, which has been extensively studied over the last few decades. The prevailing model in the field describes two layers of plant immunity (Dodds and Rathjen, 2010; Jones and Dangl, 2006). Cell surface receptors play a predominant role in the first layer of defence; located at the plasma membrane, these receptors have a primary function in preventing pathogen penetration into the cell by guarding the extracellular space. In most cases, they do so by perceiving microbe-associated molecular patterns (MAMPs) conserved across different species, such as the common bacterial ligand flagellin (Zipfel et al., 2004). Upon ligand perception, cell surface receptors, such as receptor-like kinases (RLKs) or receptor-like proteins (RLPs), trigger an immune response and prevent invasion (Gómez-Gómez and Boller, 2000; Sun et al., 2013). However, specialised pathogens have evolved to subvert this defence mechanism. They use a repertoire of small secreted proteins, known as effectors, which exploit various plant processes and promote virulence (Hogenhout et al., 2009; Win et al., 2012). A subset of effectors is translocated inside the host cell, where they carry their functions by targeting a range of host proteins and processes. In turn, they can be perceived by another class of plant immune receptors—intracellular receptors—which are the major players in the second layer of plant immunity. The majority of known cytoplasmic intracellular receptors are nucleotide-binding domain leucine-rich repeat-containing proteins (NLRs). Upon effector perception, NLRs trigger a rapid immune response, which is usually accompanied by localised programmed cell death, also known as the Hypersensitive Response (HR) (Dodds and Rathjen, 2010; Win et al., 2012). In contrast to the majority of cell surface receptors, NLRs tend to act against individual effectors that are present in a limited number of pathogen races and thus have a restricted spectrum of activity against plant pathogens. Nevertheless, the

NLR-mediated immune response is generally highly efficient and typically leads to effective restriction of pathogen proliferation.

1.2 Architecture and activities of NLR domains

NLRs are classified as STAND P-loop ATPases that belong to the AAA+ superfamily (Snider and Houry, 2008). They occur across three eukaryotic kingdoms of life, where they broadly function in stress-associated signals and immunity (Dyrka et al., 2014; Jones et al., 2016). NLR receptors share a conserved multidomain architecture that constitutes a central nucleotide-binding and oligomerisation domain (NOD) and a super-structure forming repeat (SSFR) domain (Dyrka et al., 2020). In plants, the NOD segment is exclusively formed by an NB-ARC domain (nucleotide-binding adaptor shared by APAF-1, certain *R* gene products and CED-4) followed by a leucin-rich repeat (LRR) region (McHale et al., 2006; Takken and Goverse, 2012). Much of the structural variation in NLRs resides at the N-termini; plant NLRs usually carry either Toll/interleukin 1 receptor (TIR) or coiled-coil (CC) domains at their N-termini—with the latter further divided into smaller subgroups. As a consequence, NLRs have been categorised into two major subclasses: TIR-NLRs or TNLs (TIR–NB-ARC–LRR) and CC-NLRs or CNLs (CC–NB-ARC–LRR). Representatives of both groups make up the NLR repertoire of a majority of angiosperms, although TNLs have been largely lost in monocots (Christie et al., 2016; Nandety et al., 2013; Tarr and Alexander, 2009).

Whether plant and animal NLRs share a common evolutionary origin has been under debate and there is still lack of consensus on this topic (Adachi et al., 2019b; Jones et al., 2016; Urbach and Ausubel, 2017; Yue et al., 2012). Regardless of their origin, structural conservation of NLR proteins suggests that their architecture is critical for the dual function of NLRs—non-self perception and activation of immune response. Indeed, recent elucidation of the first structures of plant NLR complexes revealed remarkable similarities to metazoan inflammasomes and apoptosomes (Hu et al., 2015; Martin et al., 2020; Wang et al., 2019; Zhang et al., 2015; Zhou et al., 2015). In a commonly accepted model, plant and vertebrate NLRs act as molecular switches that are maintained in an inactive ('off') conformation through intramolecular interactions (Takken and Goverse, 2012). In response to pathogen perception, they undergo structural remodelling that facilitates exchange from ADP to ATP at the NOD domain, leading to oligomerisation

and formation of wheel-like structures required for activation of immune response (Xiong et al., 2020). The ground-breaking study on the CC-NLR ZAR1 from *Arabidopsis thaliana* demonstrated that activated ZAR1 assembles into a higher-order complex, coined resistosome, constituting of five ZAR1 protomers (Wang et al., 2019). A similar oligomerisation mechanism has also been described for the *Nicotiana benthamiana* TIR-NLR Roq1, which forms a tetrameric resistosome upon recognition of the XopQ effector from *Xanthomonas euvesicatoria* (Martin et al., 2020). In both cases, the oligomerisation is required for cell death activation. Although there is a wealth of experimental evidence suggesting that oligomerisation is essential for activation of a number of plant NLRs (Hu et al., 2015; L. Li et al., 2020; Tran et al., 2017), whether uniform mechanisms underly activation of all plant NLRs remains to be elucidated, in particular those that engage in NLR–NLR cooperation or display atypical domain architecture. For example, the equilibrium model proposed for the L5, L6, and L7 receptors of flax predicts that these receptors exist in an equilibrium between active and inactive states, with the activated form stabilised upon recognition and thus shifting the balance towards activation (Bernoux et al., 2016). Furthermore, significant gaps in our understanding of downstream signalling processes still exit. For instance, what is the link between resistosome formation and cell death? What is the role of potential downstream signalling components in immunity activation and resistance?

1.2.1 The NB-ARC domain

The NB-ARC domain constitutes the most conserved region among distantly related NLRs, with relatively high conservation at the sequence level and perhaps even more so at the structural level (Martin et al., 2020; Xiong et al., 2020). It occupies a central position within a tripartite plant NLR and typically comprises of a nucleotide-binding domain (NBD), helical domain 1 (HD1), and a winged helical domain (WHD), which together fold into an ATP/ADP-binding NTPase (Wendler et al., 2012). The ATP and ADP molecules are coordinated and/or hydrolysed by the phosphate-binding loop (P-loop, kinase 1a, Walker A), Walker B and hhGRExE motifs within the NBD domain, the GxP (GLPL) motif of HD1, and the MHD (Met-His-Asp) motif located within the WHD subunit (Steele et al., 2019; Sukarta et al., 2016; Wang et al., 2019). All the components are highly conserved among plant NLRs and their activities have been extensively studied

in the last decades. For instance, mutations in the P-loop region significantly reduce the ATP-binding affinity, leading to complete loss-of-function phenotype in a number of NLRs, including tomato I2 receptor, N from tobacco, and more recently ZAR1 and NRG1 receptors from *Arabidopsis thaliana* (Tameling et al., 2002; Ueda et al., 2006; Wang et al., 2019). In contrast, Asp to Val (aspartic acid to valine) mutation in the MHD motif typically results in NLR constitutive activation and spontaneous cell death, as shown for M from flax and Mi from potato (Wang et al., 2015; Williams et al., 2011); this mutation is predicted to enhance ATP binding, highlighting the importance of ATPase activity in NLR activation (Sukarta et al., 2016). Notably, there is a number of exceptions, with Asp to Val mutation within the MHD motif preventing activation of Pik-1 and Pik-2 NLRs of rice (Zdrzałek et al., 2020) and other NLRs displaying P-loop–independent activity (Césari et al., 2014b; Williams et al., 2014).

Recent cryo-electron microscopy (cryo-EM) structures of an Arabidopsis CC-NLR ZAR1 (Wang et al., 2019), and a TIR-NLR Roq1 from *Nicotiana benthamiana* (Martin et al., 2020) have significantly expanded our understanding of the function of the NB-ARC domain within NLRs. They revealed that NLRs switch between an inactive ADP-bound state and an active ATP-bound state, with NB-ARC directly interacting with ADP or ATP molecules, as previously implicated (Maekawa et al., 2011; Steele et al., 2019; Williams et al., 2011). Work on ZAR1 further suggested the existence of a third, intermediate state (Wang et al., 2019). In addition, both ZAR1 and Roq1 structures demonstrated that NB-ARC plays a critical role in the structural remodelling leading to the assembly of the resistosome, as previously shown for inflammasomes and apoptosomes from metazoans (Hu et al., 2015; Zhang et al., 2015; Zhou et al., 2015). Upon activation, the NB-ARC domain undergoes drastic conformational changes; its C-terminus rotates about 180 degrees around the hinge linking the WHD and HD1 units. This rotation of the WHD subdomain, stabilised by the ATP binding, exposes the oligomerisation interface located within the NBD, enabling packing of the protomers and formation of an activated wheel-like structure. Oligomerisation mediated by the NB-ARC domain further facilitates interaction of the N-terminal CC or TIR domains, which is thought to play an essential role in activation of immune responses. Single amino acid mutations within the oligomerisation interface of the NBD prevent complex formation and thereby compromise HR (Martin et al., 2020; Wang et al., 2019).

1.2.2 The N-terminal CC and TIR domains

Plant CC-NLRs are widely present across different families of flowering plants (Lee et al., 2019). They fall into two monophyletic subclades (Prigozhin and Krasileva, 2020; Tamborski and Krasileva, 2020), each with distinct types of CC-domain: Rx-type coiled-coil with a characteristic EDVID motif (CC_{EDVID}), consisting of negatively charged residues (Collier et al., 2011; Rairdan et al., 2008); and RPW8-type CC (CC_R), found in ADR1 and NRG1 receptors (Bonardi et al., 2011; Wu et al., 2019). In addition, a new study by Lee et al. defined a third functionally conserved CC subgroup, G10-CC (CC_{G10}) (Lee et al., 2019). Whether this diversification of CC domains reflects differences in their functions and/or activities requires further investigation. Structural studies on CC_{EDVID} domains imply that they might be metamorphic and can adapt distinct topologies in response to different environmental stimuli (Bai et al., 2012; Casey et al., 2016; Hao et al., 2013). Likewise, TIR domain of Roq1 undergoes structural rearrangements upon Roq1 activation (Martin et al., 2020). Despite relatively low sequence similarity, structural topologies of TIR domains appear largely conserved across distinct plant NLRs (Bernoux et al., 2011; Williams et al., 2014; Zhang et al., 2017), suggesting that they may exhibit structural rearrangements similar to those observed for Roq1.

The N-terminal domains are thought to be associated with execution of cell death (Bentham et al., 2017). Several studies have demonstrated that TIR and CC domains are the minimal functional unit required for activation of immune response. Autoactive phenotypes have been linked to both the CC and TIR domains, with well-known examples including wheat Sr33, barley MLA10, maize Rp1-D21, Arabidopsis RPP1, and L6 from flax (Casey et al., 2016; Krasileva et al., 2010; Maekawa et al., 2011; Schreiber et al., 2016; Wang et al., 2015). However, the mechanisms underlying immune response activation for each of these domains appear to be different. Recently, some animal and plant TIR domains have been demonstrated to possess an NADase enzymatic activity (Essuman et al., 2018; Wan et al., 2019). Mutations within the putative NADase active site compromised the cell death phenotype triggered by TIR domains of Arabidopsis RPS4 and RPP1 as well as Roq1 (Martin et al., 2020; Wan et al., 2019). However, while essential, TIR NADase activity is not sufficient for activation of immune response and cell death, as illustrated by using a suite of chimeras between various TIR domains and the mammalian NLR NLRC4 (Duxbury et al., 2020). In contrast,

CC domains don't carry any known enzymatic activities. At least for a subset of them, immune response activation seems to be encoded in a functionally conserved N-terminal region of CC corresponding to the first α-helix, α1 (Adachi et al., 2019a). Transposon-based mutagenesis of NRC4 from *N. benthamiana* demonstrated that the N-terminal 29 amino acids of NRC4 are sufficient to recapitulate hypersensitive cell death. Sequence analysis paired with genome mining studies revealed that this region is defined by the consensus MADAxVSFxVxKLxxLLxxEx (MADA) motif that is conserved in nearly 20% of all CC-NLRs, including representatives from distantly related dicot and monocot species. The MADA motif is also present in the α1 of ZAR1; the helix is buried in an inactive state, yet upon activation it flips out to form a surface-exposed funnel-shaped structure in the ZAR1 oligomer (Wang et al., 2019). Mutation within this motif leads to a loss-of-HR phenotype, which is independent of oligomerisation, suggesting that it encodes other biochemical activities. Although α1 has been shown to promote plasma membrane localisation, its role there remains unknown. Finally, the remaining 80% of CC-NLRs don't carry the MADA motif, inviting the question about the alternative MADA-independent activation mechanisms (Rairdan et al., 2008; Zdrzałek et al., 2020). The majority of those likely lost the ability to execute cell death and instead rely on other NLRs to activate an immune response, as discussed in **Section 1.4.4**. However, it is still possible that a subset of CC-NLRs trigger cell death through a distinct MADA-independent mechanism.

Both TIR and CC domains contribute to the formation of NLR higher-order complexes. The structures of the TIRs from flax L6 and Arabidopsis RRS1 and RPS4 suggest that TIR–TIR protein–protein interactions are involved in TNL homo- and heterodimerization, respectively (Bernoux et al., 2011; Williams et al., 2014; Zhang et al., 2017). Experiments with plant–mammalian NLR chimeras revealed that induced proximity of the RPS4 TIR domains recapitulates hypersensitive cell death in planta (Duxbury et al., 2020). Interestingly, TIR domains within the Roq1 tetramer don't engage in tetrameric interaction and instead form only two-fold symmetric dimers, which are required for immunity activation (Martin et al., 2020). Likewise, the CC domains of the ZAR1 resistosome pack against each other resulting in a formation of an α-helical barrel (Wang et al., 2019). Yeast-two-hybrid experiments revealed heterodimerization of the CC domains of rice NLRs RGA5 and RGA4 (Césari et al. 2014b) and Pikh-1 and

Pikh-2 (Zhai et al., 2014). Notably, multiple studies demonstrated that apart from NLR–NLR interactions some TIR and CC domains are also involved in interaction with NLR-associated proteins. For instance, the CC domain of the potato Rx protein and the TIR from the tobacco N protein interact with RanGAP2 and NRIP, respectively (Mestre and Baulcombe, 2006; Sacco et al., 2007; Tameling and Baulcombe, 2007); both have been linked to NLR activation and effector perception. Future research should address the versatility of TIR and CC activities and its relevance in plant immunity as a whole.

1.2.3 The LRR domain

The leucine-rich repeat (LRR) domain is located at the C-terminal end of plant NLRs. The LRR is defined by the presence of a characteristic stretch of hydrophobic amino acids (usually leucine) spaced by hydrophilic residues, a repeat defined by a 'LxxLxxLxxNxL' motif (Wei et al., 2008). Both the length of the repeat and the overall length of the domain is highly variable (Padmanabhan et al., 2009). Despite those variabilities, the LRR domain folds into a broadly conserved structure; on the structural level LRR repeats form rigid β-sheets that act as a scaffold, in which the residues resting in between LRR units are surface-exposed (Enkhbayar et al., 2003). This characteristic feature is likely shared among a wide variety of receptors, including TIR- and CC-type NLRs in plants, animal NLR proteins, and the LRR-containing cell surface receptors, as illustrated by cryo-EM structures of Roq1, ZAR1, NLRC4, and Arabidopsis FLS2, respectively (Hu et al., 2013; Sun et al., 2013).

Multiple functions have been associated with the LRR domain. Several studies found that it mediates interaction with the SGT1-HSP90-RAR1 chaperone complex controlling stability of the inactive NLR proteins (Bieri et al., 2004; Kud et al., 2013; Leister, 2004). In cell surface receptors, the LRR is responsible for ligand binding and similar activity has also been suggested for NLR proteins (Zipfel, 2014). This is in line with the fact that surface-exposed residues in the LRR can form a large interaction interface. In addition, LRRs show very low level of conservation, indicating that they can accommodate new polymorphisms thus potentially acquiring new recognition specificities (Duxbury et al., 2016; Prigozhin and Krasileva, 2020; Van de Weyer et al., 2019). However, only few examples of effector perception through direct LRR-binding have been reported to date

(Dodds et al., 2006; Goritschnig et al., 2016; Jia et al., 2000). Studies on ZAR1 demonstrated that the LRR can also indirectly mediate effector recognition through association with pathogen-modified host proteins (Lewis et al., 2013; Wang et al., 2019). Lastly, extensive studies of NLR function indicate that the LRR domain is involved in intramolecular interactions with other domains of NLRs, which likely prevent spontaneous activation (Moffett et al., 2002; Slootweg et al., 2013; Wang et al., 2020).

1.2.4 Unconventional NLRs

Although the vast majority of all functionally validated NLRs in plants display canonical tripartite domain architectures (nearly 60% and 19% of all characterised NLRs belong to CC- or TIR-type receptors, respectively) a number of NLRs exhibits unusual domain structure (Kourelis and Kamoun, 2020). A subset of NLRs has incorporated noncanonical domains, so called integrated domains (IDs). Genome mining studies showed that integrations are much more frequent than previously thought and that NLR-IDs can represent 3–10% of all NLRs (Kroj et al., 2016; Sarris et al., 2016; Van de Weyer et al., 2019). So far, the role of integrated domains in NLR-ID fusions has only been explored for a couple of examples(Guo et al., 2018; Heidrich et al., 2013; Sarris et al., 2016), yet it is commonly thought that these noncanonical domains are involved in direct or indirect effector perception (Ellis, 2016; Wu et al., 2015).

To date, dozens of different Pfam domains have been identified in plant NLRs (Sarris et al., 2016; Steuernagel et al., 2020; Van de Weyer et al., 2019). These are associated with various functions, such as DNA binding, sugar transport, vesicle trafficking, hormone signalling, and phosphorylation (Bailey et al., 2018; Kroj et al., 2016; Sarris et al., 2016). Of these, kinases form the most prevalent group and, although their functions are yet to be experimentally determined, sequence analyses indicate that they might be catalytically active (Sarris et al., 2016). This predicted activity is in disagreement with the 'integrated decoy' model that assumes that all IDs have lost their biochemical activity and only mimic the operative targets of their cognate effectors (Césari et al. 2014a). In reality, the functions of the IDs remain to be demonstrated on a case-by-case basis.

It is also noteworthy that a subset of NLRs lack some of the canonical domains (Kourelis and Kamoun, 2020; Nandety et al., 2013; Van de Weyer et al., 2019). Some of those 'truncated' receptors have been shown to mediate pathogen recognition,

including Arabidopsis TIR-only RBA1 that recognises HopBA1 effector from *Pseudomonas syringae* (Nishimura et al., 2017) and TIR-NB protein TN 13 that is required for basal resistance (Nandety et al., 2013; Roth et al., 2017).

1.3 Pathogen perception by NLRs

Since the cloning of the first NLR genes nearly three decades ago, extensive research has revealed distinct mechanisms of effector recognition that vary tremendously across NLRs (Césari, 2018; Kourelis and van der Hoorn, 2018). Overall, NLRs can engage in two types of microbial perception either by directly binding pathogen effectors or by indirectly monitoring pathogen-induced perturbations (Win et al., 2012).

1.3.1 Direct and indirect recognition mechanisms

Physical interaction of an NLR and a pathogen effector has been proposed for several known receptors. Domain swap experiments between the flax NLRs L5 and L6 showed that N- and C-terminal regions of their leucine-rich repeat domains, respectively, are responsible for recognition specificity of alleles of flax rust effector AvrL567; direct binding to the AvrL567 alleles was confirmed in yeast two-hybrid assay and further validated in HR experiments (Ravensdale et al., 2012). Similarly, *Arabidopsis thaliana* race-specific resistance to the oomycete *Hyaloperonospora arabidopsidis* (*Hpa*, aka *Peronospora parasitica*) has been linked to association of the NLR RPP1 with different alleles of the ATR1 effector (Steinbrenner et al., 2015). In this case, only alleles that associate with RPP1 can trigger RPP1-mediated immune response (Krasileva et al., 2010; Steinbrenner et al., 2015, 2012). Despite those examples, physical association with cognate effectors has rarely been observed for plant NLRs, and indirect recognition is thought to constitute a more common recognition mechanism.

Two different models of indirect effector perception have been described in the literature: the guard model and the decoy model (van der Hoorn and Kamoun, 2008). In both cases, resistance proteins sense pathogen invasion by detecting pathogen-induced perturbation of host proteins. Functional effector targets are called guardees, while decoys are typically non-functional and only mimic effector targets. RIN4 is a negative regulator of FLS2-mediated immune response in *Arabidopsis thaliana* that constitutes one of the best

studied examples of an NLR guardee. Perhaps due to its central role in defence against bacterial pathogens, RIN4 is targeted by a number of bacterial effectors, such as AvrRpm1, AvrB, AvrRpt2, and HopF2 from *Pseudomonas syringae* (Afzal et al., 2013; Axtell and Staskawicz, 2003; Chung et al., 2011, 2014). Effector-induced modifications of RIN4 are in turn detected by the RPS2 and RPM1 NLRs leading to immunity activation and resistance. Receptor-like cytoplasmatic kinases (RLCKs), such as BIK1, belong to another class of important players in plant immunity that serve as common targets of pathogen effectors. Arabidopsis have evolved a suite of closely related pseudokinase paralogues that act as RLCK decoys cooperating in activation of immune response by the ZAR1 NLR (Lewis et al., 2013; Seto et al., 2017; Wang et al., 2015). As a consequence, ZAR1 has the ability to indirectly recognise multiple bacterial effectors from *P. syringae* and *Xanthomonas campestris*. Likewise, Prf receptor from tomato exploits activities of two unrelated effectors from *P. syringae* AvrPto and AvrPtoB. Both AvrPto and AvrPtoB bind the Pto kinase, which results in Prf-mediated immune response. Notably, Prf can interact with a suite of Pto-like kinases, some of which also prompt pathogen recognition (Gutierrez et al., 2010; Ntoukakis et al., 2014).

As illustrated by the above examples, indirect effector recognition presents a remarkable evolutionary framework, in which a single NLR can mediate resistance to multiple, often very divergent, effectors to extend its pathogen-recognition spectrum. Yet it can provide even further evolutionary advantages. In the plant–pathogen arms race, pathogen effectors constantly evolve to evade recognition, while plant NLRs adapt to those changes. Escaping indirect recognition is often linked with losing an effector or altering its activity, which might impose greater fitness costs for the pathogen than simply evading NLR binding, making indirect recognition a more effective mechanism in fighting off the invaders (Césari, 2018).

1.3.2 Recognition via integrated domain

Different still from the previously described models, some NLRs sense effectors via their integrated domains (mentioned in **Section 1.2.4**), either by binding pathogen effectors or by serving as substrates for their enzymatic activities (Césari et al., 2014a; Wu et al., 2015). The RRS1-R receptor from *A. thaliana* constitutes a well-characterised example of an NLR-ID, which detects two unrelated effectors AvrRps4 from *P. syringae*

and PopP2 from *Ralstonia solanacearum*, via its noncanonical WRKY domain near the C-terminus (Heidrich et al., 2013; Sarris et al., 2015). While recognition of AvrRps4 is mediated by direct interaction that disrupts NLR autorepression, PopP2 initiates RRS1-R activation through acetylation of WRKY (Dong et al., 2019; Heidrich et al., 2013; Sarris et al., 2015). Notably, WRKY transcription factors were shown to play an important role in plant immunity, where they upregulate expression of genes involved in plant defences upon infection (Pandey and Somssich, 2009). Therefore, using its integrated domain, RRS1-R monitors for effectors that target this family of transcription factors. Relatedly, the integrated heavy metal–associated (HMA) domains of the RGA5 and Pik-1 receptors from rice directly bind rice blast effectors AVR-Pia/AVR1-CO39 and AVR-Pik, respectively (Césari et al., 2013). Non-integrated HMAs are commonly found in a family of heavy metal–associated isoprenylated plant proteins (HIPPs) shown to contribute to abiotic and biotic stress responses (De Abreu-Neto et al., 2013; Li et al., 2020; Radakovic et al., 2018; Zschiesche et al., 2015), with rice *pi21* gene underpinning enhanced susceptibility towards *Magnaporthe oryzae* (Fukuoka et al., 2009). These observations support the hypothesis that, similar to the extraneous WRKY domain, integrated HMAs in Pik-1 and RGA5 act as baits for effector recognition. Notably, the activity of integrated domains can go beyond direct interaction with an effector. The Pii-2 receptor of rice has recently been reported to carry a NOI motif that mediates recognition of AVR-Pii, yet another effector of rice blast (Fujisaki et al., 2017). However, in contrast to previously described examples, the Pii-2 NOI domain does not physically interact with the effector itself but instead binds the host protein *Os*Exo70-F3 targeted by AVR-Pii. This mechanism of effector perception is reminiscent of indirect recognition model and shows that the function of integrated domains can differ across NLR-IDs.

Similar to the NLRs following the indirect recognition model, NLRs with extraneous domains appear to have broader recognition spectra than typical immune receptors. For example, in addition to the fact that RRS1-R WRKY recognises effectors from two distinct bacterial pathogens *Ralstonia solanacearum* and *Pseudomonas syringae*, it has also been linked to resistance against an ascomycetous fungus *Colletotrichum higginsianum* (Narusaka et al., 2013). In Pik-1 and RGA5, a single type of NLR-integrated domain binds three sequence-unrelated effectors (Césari et al., 2013; Maqbool et al. 2015; Ortiz et al., 2017). This phenomenon is likely linked with similar infection strategies deployed across

diverse pathosystems. Effectors from different pathogens have repeatedly evolved to converge on the same host signaling components and pathways, often referred to as hubs. Integration of such hubs into immune receptors might therefore increase robustness of the immune response against multiple promiscuous effectors.

1.4 Evolution of plant NLRs

Hundreds of millions of years of coevolution with pathogens have left marked footprints on plant genomes, which is particularly evident among NLR proteins (Upson et al., 2018). Advances in genome sequencing and annotation have given scientists access to a remarkable genetic diversity across various lineages of the plant kingdom, spotlighting astonishing variations that exist both between and within species but also highlighting the common patterns of NLR evolution. With the advancements in functional characterisation of plant immunity and the availability of those high-quality genomes, the mechanisms of NLR evolution and adaptation have become an emerging research topic.

1.4.1 Genomic basis of NLR evolution

1.4.1.1 Copy number variation

According to the estimates, an average number of NLR genes within plant genomes fall in a rage of 200–500 (Baggs et al., 2020), yet in reality NLR copy number varies by order of magnitude across species (Baggs et al., 2017; Christopoulou et al., 2015). Wheat and apple constitute well-known examples of species that exhibit high numbers of NLR genes, with each carrying over 1000 NLR copies (Andersen and Nepal, 2019; Shao et al., 2016; Steuernagel et al., 2020). Similarly, *Medicago truncatula* displays elevated ratio of NLRs relative to total number of proteins (Baggs et al., 2017; Sarris et al., 2016). In sharp contrast, species such as maize and members of Cucurbitaceae family have unusually low number of NLRs, with watermelon carrying as few as 31 putative functional NLRs (Baggs et al., 2017; Lin et al., 2013; Wu et al., 2015). While the basis of sparsity of NLR genes in maize remain to be elucidated, studies in Cucurbitaceae suggest that low NLR copy numbers in this family are likely a consequence of frequent losses and sparse duplication events (Lin et al., 2013). Relatedly, Baggs and colleagues have recently shown

that many aquatic species underwent independent loss of NLR genes, including humped bladderwort (*Utricularia gibba*) encoding at most one NLR (Baggs et al., 2020).

In addition to interspecies variations, NLRs are known to display frequent intraspecies presence/absence (P/A) polymorphisms, exemplified by polymorphisms in lettuce (Christopoulou et al., 2015; Kuang et al., 2004) and Arabidopsis (Henk et al., 1999; Prigozhin and Krasileva, 2020; Shen et al., 2006). In rice, 124 NLRs were shown to exhibit P/A polymorphism between two different cultivars 93-11 and Nipponbare (Luo et al., 2012). Likely, ever-expanding genomic datasets will provide further and more comprehensive insights into intraspecific NLR diversity, as illustrated by the recently published pan-NLRome of Arabidopsis (Van de Weyer et al., 2019).

The high turnover of NLRs has been conceptualised through the birth-and-death model (Michelmore and Meyers, 1998), explaining linage-specific expansions and/or contractions of NLR genes. Although the exact selection mechanisms driving these processes remain elusive, the expansions are usually associated with selection pressure imposed by plant pathogens, which favours beneficial NLRs with altered specificity. Indeed, the levels of diversifying selection and rates of P/A polymorphisms broadly correlate with NLR function, with NLRs not involved in pathogen perception showing the highest conservation rate (Prigozhin and Krasileva, 2020; Stam et al., 2019; Van de Weyer et al., 2019; Wu et al., 2017). Moreover, differences in effector repertoires in indica- and japonica-borne isolates of rice blast correlate with considerable differences in immunity between indica and japonica rice varieties (Liao et al., 2016; Xiahong et al., 2011). The NLR repertoires can also reflect plant lifestyle and adaptation to the environment, as illustrated by the convergent reduction in the numbers of NLRs in aquatic species (Baggs et al., 2020) or different patterns of within-species expansions in response to geographical regions in wild tomato (Stam et al., 2016); to date, however, the evolutionary mechanisms driving these differences are poorly understood. Finally, plants often lose unutilised NLRs, which could be explained by either relaxed or purifying selection. Surely, the evolutionary drivers of contractions of NLR repertoires should be considered on a case-by-case basis but there is substantial evidence indicating that NLRs can incur fitness costs (Hulbert et al., 2001; Karasov et al., 2014; MacQueen et al., 2016) and experience high levels of pseudogenisation (Lin et al., 2013; Meyers et al., 2003;

Xu et al., 2011), probably owing to genetic incompatibilities or energetic costs associated with maintenance of unused genes (Karasov et al., 2014; MacQueen et al., 2016).

1.4.1.2 Genomic localisation and organisation

NLR genes exhibit uneven distribution, with some existing as physical singletons and others located in clusters that vary significantly in structure and size (Meyers et al., 1998; Stam et al., 2016; van Wersch and Li, 2019; Zhang et al., 2016). For example, in lettuce the vast majority of all NLR genes lie within five major regions (Christopoulou et al., 2015), whereas in Nipponbare rice, the chromosome 11 alone encodes nearly a quarter of all NLRs (Zhou et al., 2004). In addition, there are significant variations at the level of single species; in Arabidopsis, clusters comprise of 47 to 71% of all NLRs, depending on the accession (Van de Weyer et al., 2019). Interestingly, the genetic make-up of such clusters varies greatly across and within species. On average around half of them appears to be heterogeneous while the other half consists primarily of phylogenetically related genes (homogeneous) (Zhang et al., 2016), with examples of both found in various plant species (Christopoulou et al., 2015; Holub, 2001; Jupe et al., 2012; Leister et al., 2005; Zhang et al., 2016). While the emergence of homogeneous clusters is usually attributed to tandem duplication, the origin of heterogeneous blocks is less apparent and could involve ancient duplications and chromosomal rearrangements (Hulbert et al., 2001).

This unique genomic organisation likely accelerates NLR evolution and favours gene diversification. Additional gene copies might provide a larger mutational target but also allow relaxed selection among gene paralogues, so they can diversify and act as a reservoir of genetic diversity. Indeed, clustered genes tend to display higher nucleotide diversity indicative of relaxed selection (Prigozhin and Krasileva, 2020; Van de Weyer et al., 2019). For example, the highly homologous Rx and Gpa2 receptors of potato evolved to confer resistance to two distinct intruders, potato virus X and potato cyst nematode *Globodera pallida*, respectively (Van Der Vossen et al., 2000). Although this model appears to explain much of NLR genetic diversity, there are examples of physical singletons exhibiting relatively high diversity and clustered NLRs with high conservation levels (Prigozhin and Krasileva, 2020). Finally, gene clusters are often associated with an increased recombination rate (Hulbert et al., 2001; Kuang et al., 2004; Meyers et al., 1998;

Wicker et al., 2007), though some clusters can also experience supressed recombination (Jiao and Schneeberger, 2020).

1.4.1.3 Mechanisms of NLR expansion and diversification

Ectopic recombination, unequal crossing over, and transposition are among the major mechanisms underlying expansion of NLR genes, with point mutations, gene fusions, and gene conversion further contributing to NLR diversification (Bailey et al., 2018; Upson et al., 2018; Wicker et al., 2007). Comparative analysis of three pepper genomes and other members of Solanaceae family revealed that retrotransposition may have facilitated NLR expansion in Solanaceae (Kim et al., 2017a). Tomato *I2*, potato *R3a*, and pepper *L* genes constitute examples of well-known NLRs that may have originated from a single gene by retrotransposition followed by subsequent neofunctionalisation in each plant lineage. Given that NLR genes are frequently located in close proximity to different types of transposon elements (Christopoulou et al., 2015; Golicz et al., 2016; Henk et al., 1999), and the number of NLR clusters positively correlates with density of transposon elements on the same chromosome (Li et al., 2010), it is possible that transposon-driven duplications can be common in plants (Seidl and Thomma, 2017; Wicker et al., 2010).

As previously described (**Section 1.4.1.2**), high similarity clusters emerge through tandem duplications and functional diversification, which can be driven by unequal crossing-overs, intra-cluster rearrangements, or gene conversion (Kuang et al., 2004; McDowell et al., 1998; Van de Weyer et al., 2019; Wicker et al., 2007). For instance, the *MRC1* locus of lettuce consists of 60 out of a total of 62 *RGC16* NLR family members, which likely emerged through illegitimate recombination (Christopoulou et al., 2015). Another remarkable example of recent duplication within an NLR cluster is *Rp1-D* of maize. *Rp1-D* is a complex locus encompassing nine NLR paralogues—each with over 90% nucleotide sequence identity—that arose through a series of unequal crossing-overs (Smith et al., 2010). The genes within the *Rp1* locus often mis-pair in meiosis, resulting in novel gene combinations or, occasionally, novel genes.

1.4.2 NLR diversity at individual loci

Both plants and pathogens are under strong selection pressure and their coevolutionary dynamics are often defined by an evolutionary arms race. As a consequence, immune receptors carry strong signatures of diversifying selection (Seeholzer et al., 2010; Stam et al., 2019; Van de Weyer et al., 2019) as evidenced by high allelic diversity. One example comes from the flax *L* gene that encodes 13 alleles of the L NLR (Dodds et al., 2006; Ellis et al., 1999). Despite over 90% sequence similarity, L proteins display distinct recognition specificities, mediating race-specific resistance to the rust fungus *Melampsora lini*. Functional studies revealed that pathogen recognition spectra can be traced to antagonistic molecular interactions with the cognate AvrL567 effector alleles of *M. lini* (Dodds et al., 2006; Ravensdale et al., 2012). Similarly, Arabidopsis RPP13 is involved in coevolutionary arms race with the ATR13 effector of *Hyaloperonospora arabidopsidis* (Allen et al., 2008; Hall et al., 2009). Both the NLR and the effector in the RPP13–ATR13 interaction exhibit high allelic diversity (Allen et al., 2008; Krasileva et al., 2011). Notably, *RPP13* is located in a crossover hotspot, which have most likely accelerated RPP13 evolution, making it the most polymorphic single-gene NLR known to date (Prigozhin and Krasileva, 2020; Serra et al., 2018).

Although NLR alleles are frequently functionally related, many alleles neofunctionalise to recognise sequence-unrelated effectors. *Mla* is one of best studied multi-functional loci that confers resistance to barley powdery mildew fungus, *Blumeria graminis* f. sp. *hordei* (Maekawa et al., 2019; Wei et al., 2002). Extensive research on the MLA receptors revealed that they can perceive a suite of unrelated powdery mildew effectors, collectively named the AVR_a effectors (Lu et al., 2016; Saur et al., 2019). Interestingly, the *Rmo1* gene conferring resistance to *M. oryzae* was also mapped to the *Mla* locus, yet the molecular mechanism underpinning *Rmo1*-mediated blast resistance remains unknown (Inukai et al., 2006). Furthermore, wheat orthologues of *Mla Sr33* (Periyannan et al., 2013) and *Sr50* (Mago et al., 2015) confer resistance to the Ug99 isolate of the stem rust pathogen *Puccinia graminis* f. sp. *tritici*, indicating that *Mla* ancestor evolved to recognise distinct pathogen effectors from unrelated fungi.

While the arms race model seems to explain the evolutionary dynamics between a single NLR and its cognate effector, it is often insufficient to capture the full complexity of plant and pathogen populations and their relationships (Bergelson et al., 2001; Meyers

et al., 2005). In reality, many NLRs and disease-related genes are maintained though long-term balancing selection that could reflect fitness costs associated with resistance, local adaptations, population fluctuations, or diffuse community-wide interactions (Bakker et al., 2006; Brown and Tellier, 2011; Karasov et al., 2014; Koenig et al., 2019; MacQueen et al., 2016; Sela et al., 2014). Brabham et al. have recently shown that a member of the *Mla* locus *RGH2* was subjected to fusions at its 3' end (Brabham et al., 2017). Although the majority of RGH2 homologues carry an integrated Exo70 domain, there is large variation in domain structure among RGH2 homologues, including presence/absence of integration, which may be indicative of balancing selection acting on the gene. Intraspecific maintenance of polymorphisms was also shown for RPS5 (Karasov et al., 2014) and RPS2 (MacQueen et al., 2016) from *Arabidopsis thaliana*.

1.4.3 Gene fusions and integrations

The acquisition of extraneous domains in aforementioned NLR-IDs (**Sections 1.2.4** and **1.3.2**) is a remarkable example illustrating how pathogens shape evolution of plant immune receptors. Given that several integrated domains show homology to signalling components required for immune responses, they are generally thought to have derived from effector operative targets, which then act as baits for effector recognition within NLRs (Césari et al., 2014a; Wu et al., 2015). Therefore, NLR-ID fusions provide marked evolutionary advantages against promiscuous effectors that often converge on the same target, with the astonishing number and diversity of integrated domains pointing to how successful this evolutionary strategy is (Upson et al., 2018). Nevertheless, the frequency and mechanisms underpinning the emergence of NLR-ID fusions, as well as the degree to which novel fusions are maintained in plant populations, remain poorly understood.

In search for answers, Bailey et al. analysed NLR diversity in Poaceae (Bailey et al., 2018). They discovered a 'major integration clade', which includes *RGA5, Pi-ta*, and *Rpg5*, that underwent repeated independent integration events; existence of the major integration clades was subsequently validated in wheat (Steuernagel et al., 2020). Based on this observation, the authors proposed that ectopic recombination is a major driver of domain integration in grasses and that some NLR loci may be located in hotspots of recombination that favour the integration. Further analysis of wheat and barley NLR-IDs suggests that the fusions are regulated by alternative splicing, with certain genes

encoding transcripts with or without the integration (Andersen and Nepal, 2019). Indeed, the well-characterised *RGA5* gene was also shown to encode two alternative transcripts, one including and one excluding the HMA domain (Césari et al., 2013).

Notably, not all integrated domains appear to mediate pathogen perception. The yellow rust resistance genes Y*r4*, *Yr5*, and *YrSP* (Marchal et al., 2018), as well as rice *Xa1* and *Xo1* that mediate resistance to *M. oryzae* and bacterial blight or bacterial leaf streak, respectively (Read et al., 2020; Yoshimura et al., 1998), belong to a group of evolutionary-related genes, which reside in one of the major integration clades described by Bailey et al. (Bailey et al., 2018). These NLRs share similar domain architecture, with integrated zinc-finger BED (zfBED) domain located towards the N-termini of the proteins. While *Yr* fusions appear to have derived from a single integration, it is unclear whether their rice counterparts (and their relatives) acquired the zfBED domain through an independent integration event. In addition, zfBED-containing NLRs have also been detected in dicot species, namely poplar (*Populus trichocarpa*) (Germain and Séguin, 2011). The zfBED-integration presents an interesting case, in which the integrated domain does not appear to govern pathogen recognition specificity (Marchal et al., 2018; Read, 2020), inviting the question about the role of zfBED within those NLRs. Given that this domain belongs to a family of DNA-binding proteins (Aravind, 2000) and that Xa1 and Xo1 appear to carry a nucleus-localisation signal (NLS) (Read et al., 2020), it's interesting to speculate that this extraneous domain might play a specialised role in activation of immune signalling through interaction with DNA.

1.4.4 Functional diversification—singletons, gene pairs, and networks

The gene-for-gene model postulated by Harold Henry Flor in the middle of the 20th century proposed that every plant resistance gene has a matching pathogen counterpart (Flor, 1971). Although some NLRs indeed appear to function as singletons, thus following Flor's one-for-one model, in many cases the underlying genetic architecture of disease resistance is much more complex and is not limited to a single resistance gene (Adachi et al., 2019b).

Typically, NLRs are considered singletons if they can recognise matching effector and trigger cell death in a distantly related heterologous plant system. The members of the MLA family are some of best-known functional singletons. The MLA alleles are

involved in direct interaction with powdery mildew AVR_a effectors and can trigger immune response in the model dicot plant *Nicotiana benthamiana*, suggesting that they function as singletons (Maekawa et al., 2019; Saur et al., 2019; Wei et al., 2002). The RPS5 NLR from Arabidopsis is another likely singleton. Co-expression of RPS5 with the PBS1 kinase and the cognate effector from *P. syringae*, AvrPphB, is sufficient to trigger RPS5-mediated HR in *N. benthamiana* (Ade et al., 2007; Qi et al., 2012). Finally, reconstitution of ZAR1 and ROQ1 resistosomes provide substantial evidence that they that don't require NLR signalling partners for activation of immune response (Martin et al., 2020; Wang et al., 2019).

Some NLRs function in pairs. In a paired model, a sensor NLR that specialises in perceiving pathogen-derived effectors requires a helper NLR (also referred to as an executor NLR) to activate immune signalling (Bonardi et al., 2011; Eitas and Dangl, 2010). Helper/sensor pairs are typically encoded by two genetically linked genes, which is thought to facilitate gene co-segregation and transcriptional co-regulation, ensuring effective pathogen recognition and signal transduction (Griebel et al., 2014; Krom and Ramakrishna, 2008). Interestingly, all well-characterised NLR-IDs are functionally linked, including aforementioned RPS4, RGA5, Pik-1, and Pii-2 (**Section 1.4.3**) that rely on RRS1, RGA4, Pik-2, and Pii-1 helper NLRs, respectively (Césari et al., 2013; Dong et al., 2019; Fujisaki et al., 2017; Kanzaki et al., 2012; Narusaka et al., 2009; Zdrzałek et al., 2020). It is possible that functionally specialised paired NLRs have an increased tolerance to integration of noncanonical domains given that the ID is less likely to perturb the function of an NLR pair than of a singleton constrained by its signalling activity (Upson et al., 2018).

The Pia (RGA4 and RGA5) and the Pik pairs of rice illustrate the two distinct mechanisms by which helper/sensor pairs work. They are both encoded by single loci in head-to-head orientation, and both confer resistance to *M. oryzae* by recognising effectors via the sensor-integrated HMA domain (Césari et al., 2014b, 2013; De la Concepcion et al., 2020, 2018; Kanzaki et al., 2012; Maqbool et al., 2015; Okuyama et al., 2011; Ortiz et al., 2017; Yuan et al., 2011). Despite these similarities, the mechanisms by which they work differ. In the Pia pair, the RGA4 helper is constitutively active and triggers spontaneous cell death when expressed in the absence of its partner, RGA5, that acts as a negative regulator (Césari et al. 2014b). Direct binding of the blast

effector AVR-Pia or AVR1-CO39 to the RGA5 receptor results in the release of RGA4 from negative regulation and activation of immune signalling. In contrast, neither Pik-1 nor Pik-2 trigger spontaneous cell death in the absence of pathogen stimuli (Kanzaki et al., 2012; Maqbool et al., 2015; Zdrzałek et al., 2020). Instead, they function though fine-tuned cooperation, in which both partners are required for signalling (Zdrzałek et al., 2020).

In addition to pairs, NLR proteins can also function as sophisticated signalling networks. Recently, nearly a third of all CC-NLRs of Solanaceae was shown to comprise the NRC network, in which a large number of sensor NLRs rely on few helpers, the NRCs, to activate immune response (Wu et al., 2017, 2016). While NRCs carry a largely conserved MADA motif, the sensors appear to have degenerated at their N-termini, which could be indicative of evolution towards effector perception or relaxed selection (Adachi et al., 2019a). Phylogenetically, members of the NRC network fall into a well-supported superclade that most likely evolved from a single sensor/helper pair nearly 100 million years ago (MYA) (Wu et al., 2017). Another immunoreceptor network is defined by the helper NLRs belonging to the ADR1 or NRG1 gene families that function downstream of a diverse array of sensor NLRs, both TIR- and CC-NLRs (Bonardi et al., 2011; Castel et al., 2019; Jubic et al., 2019; Qi et al., 2018; Wu et al., 2019). Interestingly, ADR1 and NRG1—which both belong to the CC_R-NLR clade—emerged before the split of gymnosperms and angiosperms, suggesting that the NRG1/ADR1 network constitutes an ancient signalling mechanism that spans across diverse plant families (Shao et al., 2016; Stam et al., 2019). The NRC and NRG1/ADR1 networks likely differ at the mechanistic level, yet from the evolutionary standpoint they share remarkable similarities, in which a diverse set of sensors that mediate resistance to pathogens across kingdoms converge on few helper hubs (Adachi et al., 2019b; Castel et al., 2019; Wu et al., 2017). In both cases, the sensors underwent massive expansions and functional diversifications while the helpers diversified at a much slower pace, possibly due to constraints imposed by the need for signalling activity (Prigozhin and Krasileva, 2020; Shao et al., 2016; Stam et al., 2019). Potentially, the classic bow-tie architecture of these signalling networks may be the consequence of massive sensor NLR diversification caused by coevolution with rapidly evolving pathogens, and is reminiscent of the

networks of plant cell-surface immune receptors and animal Toll-like receptors (Oda and Kitano, 2006; Smakowska-Luzan et al., 2018).

To conclude, the requirement of an NLR signalling partner by ligand-sensing receptors, exemplified by NLR pairs and networks, appears to be a recurrent theme in NLR evolution. Adachi et al. proposed that the functional specialisation has been a crucial evolutionary event that enhanced NLR evolvability (Adachi et al., 2019b). NLRs involved in helper–sensor relationships tend to display asymmetrical evolution. While helpers stay relatively conserved, sensors experience elevated rates of diversification, raising a possibility that they become solely devoted to effector perception and thereby lose their signalling capacity. This conceptual framework coupled with functional characterisation and computational pipelines could help answer open questions about the mechanisms of activation and interconnections within plant NLRomes. What is the molecular basis of the transition from singletons to pairs and networks? Are there any structural features that define helpers and sensors and underpin helper/sensor specificity? What are the mechanistic overlaps and differences among NLR pairs and networks? Can we leverage sequence and phylogenetic information to predict NLR modes-of-action and discover novel NLR connections?

1.4.5 NLR coevolution and autoimmunity

Evolution of disease resistance is primarily, and rightly so, associated with adaptation towards pathogen recognition. However, biological systems evolve simultaneously at the level of sequence, function, and structure, with different selection pressures acting upon them (Bastolla et al., 2017; Dean and Thornton, 2007). For instance, physical properties are an integral part of the evolution of biological molecules with conformation and stability manifested in a number of ways in protein sequences and structures. NLRs are not exempt from the laws of physics and chemistry and evolve accordingly. Perhaps the biggest evolutionary constraint for NLRs is the nature of their activities; given that NLR function is associated with cell death, there is a high selection pressure to keep the protein in an inactive state in the absence of pathogen stimuli. Ultimately, NLR-triggered immunity is a fine-tuned process and even the slightest manipulation in NLR structure or sequence can lead to uncontrolled spontaneous activation of the immune response (Alcázar et al., 2009; Bomblies et al., 2007; Bomblies and Weigel, 2007; Chae et al., 2016;

Deng et al., 2019; Yamamoto et al., 2010). Overexpression of some NLRs can also result in autoimmunity, thus NLR expression is thought to be tightly regulated through chromatin modifications or transcription factors (Heidrich et al., 2013; Lai and Eulgem, 2018). Moreover, recent studies have emphasised the role of microRNAs in negative regulation of NLR gene expression (De Vries et al., 2015; Zhang et al., 2016). Lastly, NLRs don't evolve in isolation—they function in the context of the plant and coevolve with other plant components (Baggs et al., 2020).

1.4.5.1 Intramolecular interactions

Functional studies of closely related genes have helped to dissect the intramolecular interactions constraining NLR evolution. One example comes from potato, which carries two highly similar homologues Rx1 and Gpa2 that diversified to confer resistance against the potato virus X and the potato cyst nematode, *Globodera pallida*, respectively (Van Der Vossen et al., 2000). Sequence exchange between these homologues revealed that mismatch between the WHD domain and the N-terminal part of the LRR results in loss-of-activity or spontaneous cell death (Rairdan and Moffett, 2006; Slootweg et al., 2013). Nonetheless, both Rx1 and Gpa2 have retained their capacity to function with their NRC mates (Wu et al., 2017). A similar observation was made for *Rp1-D21*, which is derived from an intragenic recombination between *Rp1-D* and *Rp1-dp2* (Smith et al., 2010; Smith and Hulbert, 2005; Wang et al., 2015). In contrast to the 'parental' genes, *Rp1-D21* causes a 'lesion mimic' phenotype manifested by spontaneous HR, which is further enhanced by pathogen inoculation. Both examples highlight functional and biophysical constraints shaping the intramolecular evolution of plant NLRs and the importance of coordinated sequence changes—coevolution—across NLR structure and sequence.

1.4.5.2 Genetic incompatibility

NLRs and other immunity-related proteins are frequently involved in hybrid necrosis, also known as Bateson-Dobzhansky-Muller-type incompatibility, in which epistatic interactions of two or more loci can have deleterious effects on plant fitness (Alcázar et al., 2009; Bomblies et al., 2007; Bomblies and Weigel, 2007; Chae et al., 2016;

Deng et al., 2019; Yamamoto et al., 2010). Perhaps the most prominent examples of hybrid incompatibilities have come from a systemic hybrid screen of Arabidopsis accessions that yielded over a hundred cases of necrosis (Chae et al., 2014). Remarkably, many of the incompatibilities were repeatedly linked to highly polymorphic NLR clusters, including well-known *RPP* resistance loci encoding genes mediating disease resistance to *H. arabidopsidis* (Bomblies et al., 2007; Chae et al., 2014; Prigozhin and Krasileva, 2020). Follow-up studies revealed that spontaneous cell death phenotype for a number of hybrids was due to incompatible intermolecular interactions between divergent alleles of NLRs or other disease-related molecules, such as RPW8-like proteins (Barragan et al., 2019; Li et al., 2020). Furthermore, biochemical analysis of the mismatch between the RPP7 and HR4 proteins linked the autoimmunity to NLR activation and resistosome formation; in the proposed model, HR4 interferes with RPP7 self-inhibition leading to formation of higher-order complex and cell death (Li et al., 2020). Although, there is some evidence indicating that certain incompatible alleles may exist in geographically isolated populations, thereby reducing the risk of lethal hybrids, there are also cases where such alleles coexist at low frequencies in the same population (Barragan et al., 2020; Todesco et al., 2014).

Noteworthy, gene incompatibilities are not restricted to necrosis and can sometimes result in suppression of resistance, which is a common phenomenon in plant breeding (Hurni et al., 2014). Introgression of rye *Pm8* gene is ineffective in conferring race-specific resistance against wheat powdery mildew, *B. graminis*, in wheat lines that carry *Pm8* homologue *Pm3*. The repression was linked to the formation of a PM3–PM8 heterocomplex, which presumably prevents activation of PM8. In Arabidopsis, inappropriate sensor/helper combinations of homologous NLR pairs RRS1/RPS4 and RRS1B/RRS4B fail to trigger immune response, possibly as a result of weaker association between mismatched partners (Saucet et al., 2015). It is remarkable that NLR genetic incompatibility has been observed within populations of the same species or, like in the case of RRS1/RPS4, in the same genome, highlighting the extremely dynamic nature of NLR evolution.

1.5 Mechanistic approaches to the study of evolution of plant–microbe interactions

Plant–microbe systems are remarkable in their evolutionary dynamics that can be studied across multiple timescales (Upson et al., 2018). Nonetheless, the mechanistic research in the field of plant–microbe interaction is often conducted without a robust phylogenetic and ecological framework and without the appreciation of how those systems came to be. Conversely, evolutionary research often fails to integrate the incredible range of molecular mechanisms and models that continue to emerge in the field. As a consequence, interdisciplinary studies linking mechanistic molecular and evolutionary approaches have rarely been conducted, but the trend *is* beginning to change.

Questions about biochemical drivers of adaptation can be addressed by reconstructing the evolutionary trajectories of the proteins of interest (Dean and Thornton, 2007; Harms and Thornton, 2013). Using phylogenetic techniques and algorithms for ancestral sequence reconstruction it is now possible to statistically infer ancestral sequences, which can then be synthesized, expressed, and experimentally studied in the context of modern sequences (**Figure 1.1**) (Ashkenazy et al., 2012; Cohen and Pupko, 2011; Pupko et al., 2000). Ancestral sequence reconstruction has been particularly fruitful in enzymology, providing insights into the mechanisms of functional divergence and molecular adaptation (Nguyen et al., 2017; Parrent et al., 2009). In the field of plant–microbe interactions, resurrection of the cystatin-like protease inhibitor EPIC1 has helped deepen our understanding of effector specialisation and adaptive evolution of the oomycete pathogen *Phytophthora mirabilis* following a host jump (Dong et al., 2014). More recently, characterisation of the extinct Tin2 effector was used to challenge hypothetical models of its evolution, revealing that the function of the present-day *Ustilago maydis* Tin2, which appears to contribute to the pathogenic lifestyle of this species, is evolutionarily derived (Tanaka et al., 2019; Zess et al., 2019). Both studies have shown that experimental analyses of an ancestral effector can transcend phylogenetic inference to yield more accurate evolutionary models (Dong et al., 2014; Tanaka et al., 2019; Zess et al., 2019).

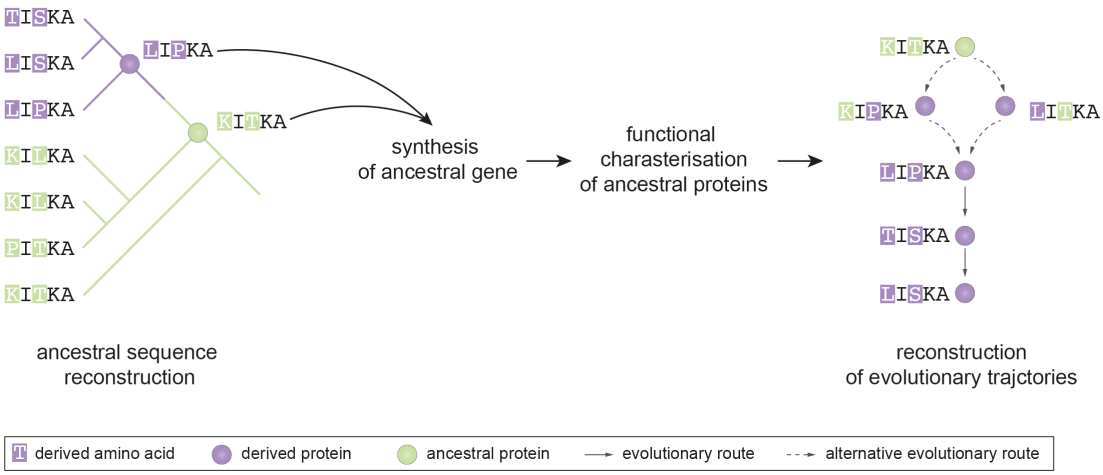


Figure 1.1 Schematic of ancestral protein reconstruction

Statistical approximations of ancestral sequences at the internal nodes of the phylogenetic tree are computationally inferred based on sequence alignment and phylogenetic relationship of modern sequences. Ancestral proteins can be 'resurrected' by synthesising the genes encoding predicted proteins. These resurrected proteins can then be functionally characterised, for instance by using biochemistry methods. By means of reconstructing the evolutionary trajectory of a sequence and characterising ancestors from different nodes it is possible to correlate shifts in protein phenotype with a specific genotype and time interval. Figure adapted from: Harms and Thornton, 2013.

1.6 Effectors of plant pathogens

Plant pathogen effectors are small secreted molecules that modulate host processes to promote colonisation and pathogen growth (Franceschetti et al., 2017; Hogenhout et al., 2009). Despite being encoded by pathogen genomes, effectors function in the context of the plant and exhibit marked adaptations towards plant proteins and physiological processes. They carry an incredible diversity of functions, which include but are not limited to, suppression of plant immunity, inhibition of plant enzymes, modulating gene expression, and rerouting of nutrients (Win et al., 2012). Some exhibit enzymatic activity, while others bind host proteins to modulate their activities. Some are translocated inside the plant cell and others function in the apoplastic space.

Effectors frequently converge on the same host pathways, whether they originate from a single pathogen or from phylogenetically unrelated ones. An example of functional redundancy comes from *P. syringae*, where four phylogenetically unrelated effectors deploy distinct strategies to target the same Arabidopsis protein RIN4 (Grant et al., 2006). Other examples include AVR-Pik, AVR-Pia, and AVR1-CO39 from *M. oryzae*. Each of these

effectors bind rice proteins with an HMA domain (Césari et al., 2013; De la Concepcion et al., 2020, 2018; Guo et al., 2018; Kanzaki et al., 2012; Maqbool et al., 2015; Ortiz et al., 2017). Although the precise identity of their host targets remains obscure, all three are thought to bind HMA-containing proteins in order to promote infection. Similarly, two bacterial effectors AvrRps4 and PopP2 appear to target Arabidopsis WRKY transcription factors, even though they derive from different bacterial pathogens, *P. syringae* and *R. solanacearum*, respectively (Heidrich et al., 2013; Sarris et al., 2015).

One of the hallmarks of pathogen effectors is that they tend to share little or no sequence similarity to each other or to other known proteins. The three-dimensional structures of effector proteins have, however, revealed unexpected similarities between phylogenetically unrelated effectors (Franceschetti et al., 2017). One example comes from members of the RXLR effector family from oomycetes that carry a characteristic fold defined by consensus sequence motifs W, Y, and L (Boutemy et al., 2011; He et al., 2019; Win et al., 2012). These WY- or LWY-domains can occur as single units or in tandem repeats (He et al., 2019; Raffaele and Kamoun, 2012). In *Phytophthora* spp. around 44% of all RXLR effectors contain at least one WY-domain (Boutemy et al., 2011). The aforementioned AVR-Pik, AVR-Pia, and AVR1-CO39 as well as AvrPiz-t from *M. oryzae* belong to another structurally conserved class of effectors, termed MAX effectors (for *Magnaporthe* Avrs and ToxB-like) (de Guillen et al., 2015). MAX proteins fold into a six-stranded β-sandwich with two antiparallel β-sheets, which form an extended surface-exposed area. In *M. oryzae* and in the related species *M. grisea*, between 5 to 10% of the entire effector repertoire is predicted to adapt a MAX fold (de Guillen et al., 2015; Petit-Houdenot et al., 2020). It is remarkable that the same structure is also shared with the host-selective toxin ToxB from *Pyrenophora tritici-repentis*, despite being derived from a phylogenetically distant pathogen and having no apparent sequence similarities (de Guillen et al., 2015). de Guillen et al. proposed that MAX proteins likely expanded through diversifying rather than convergent evolution, inviting the possibility that this unique fold provides a flexible platform to evolve different activities, while maintaining structural integrity.

Plant hosts are continuously shaping pathogen evolution, driving recurrent and sustained changes within microbial genomes that are particularly evident in effector repertoires (Upson et al., 2018). Effectors are under dual selection pressure: the pressure

to adapt to host targets and the pressure to evade detection by plant immune receptors. As a result, they carry extreme signatures of positive selection, manifested by high rates of nonsynonymous substitutions relative to synonymous changes (Allen et al., 2004; Dodds et al., 2006; Huang et al., 2014; Raffaele et al., 2010). Positively selected sites often map to regions that underpin effector's activity or surface-exposed residues mediating interaction with immune receptors (Huang et al., 2014; Liu et al., 2005; Yoshida et al., 2009). In addition to high sequence divergence levels, pathogen effectors experience frequent presence/absence polymorphisms (Fouché et al., 2018). Effector repertoires have also been linked to host specificity, with many well-documented examples in *Magnaporthe oryzae*, including PWL effector family (Kang et al., 1995) and more recently PWT3 and PWT4 effectors, whose losses facilitated the host jump from *Lolium* to wheat and subsequent emergence of wheat blast (Inoue et al., 2017).

1.7 Rice–rice blast interaction as a model system for studying NLR–effector coevolution

The ascomycete fungus *Magnaporthe oryzae* (anamorph *Pyricularia oryzae*) is a causal agent of blast disease of cereals. The host range of *M. oryzae* spans across two major grass lineages: the PACMAD clade (for Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, and Danthonioideae) comprising of maize and sorghum, and the BOP clade (for Bambusoideae, Oryzoideae, and Pooideae) that includes agronomically important crops, such as rice, wheat, and barley. In total, blast fungus infects nearly 50 different grass species and can cause 10–30% yield losses annually (Talbot, 2003). Despite being a multihost pathogen, *M. oryzae* forms genetically distinct lineages that tend to be specialised towards a limited number of hosts (Gladieux et al., 2018). Rice-infecting isolates belong to a monophyletic lineage of *M. oryzae* that is thought to have originated from isolates infecting foxtail millet (*Setaria* spp.) following a host jump dated at around 7,000–9,000 years ago (Couch et al., 2005; Gladieux et al., 2018; Latorre et al., 2020; Yoshida et al., 2016). This jump to domesticated rice is commonly attributed to extensive gains and losses of genes associated with transposable elements and nucleotide polymorphisms (Yoshida et al., 2016).

Notably, effectors are encoded by genes that most frequently exhibit presence/absence polymorphisms or elevated rates of nonsynonymous substitutions relative to synonymous mutations attributed to host specialisation (Yoshida et al., 2016). One of the best-characterized examples is the AVR-Pik effector. AVR-Pik is widely present across different rice-infecting isolates, but not in other *M. oryzae* lineages, suggesting the possibility that it might have contributed to specialization of *M. oryzae* on rice (Langner et al., 2020; Latorre et al., 2020; Yoshida et al., 2016). In addition, AVR-Pik is a striking example of effector adaptation imposed by an NLR. Allelic variants of AVR-Pik carry only five amino acid replacements, but no synonymous changes (Huang et al., 2014; Kanzaki et al., 2012; Longya et al., 2019; Yoshida et al., 2009). It is astonishing that all nonsynonymous mutations map to regions in the protein structure that are located at the binding interface with the cognate immune receptor Pik-1, indicating the adaptive nature of those polymorphisms (De la Concepcion et al., 2020, 2018; Longya et al., 2019; Maqbool et al., 2015). Likewise, the integrated HMA domain responsible for effector recognition is the most diversified region among Pik-1 allelic variants, consistent with the view that the receptor is under strong selection imposed by the AVR-Pik effector (Białas et al., 2018; Costanzo and Jia, 2010; De la Concepcion et al., 2020; Zhai et al., 2014). Interestingly, the most ancient of the AVR-Pik alleles, AVR-PikD, is recognised by all Pik-1 variants characterised to date, whereas the most recent effector allele, AVR-PikC, evades recognition by all known Pik-1 alleles, possibly reflecting the coevolutionary arms race between these *M. oryzea* effectors and rice NLRs (**Figure 1.2**) (Białas et al., 2018; Kanzaki et al., 2012). This work focused on using the Pik-1–AVR-Pik interaction as a model system for studying coevolution between plants and plant pathogens and provide a broader understanding of the mechanisms of adaptation and diversification of plant NLRs across multiple timescales.

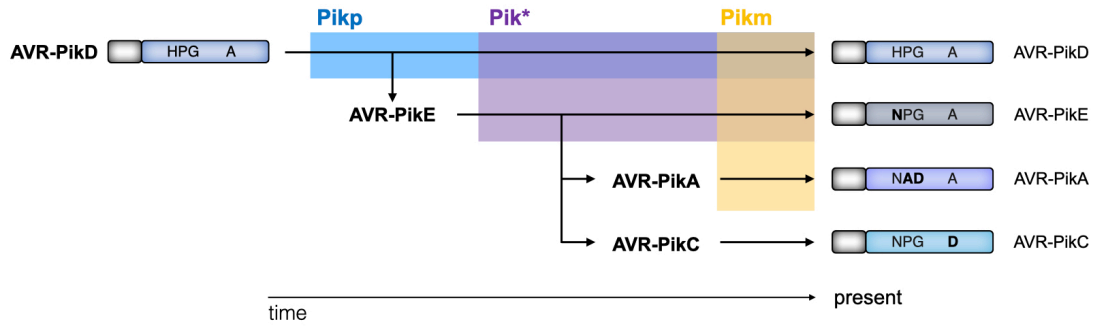


Figure 1.2 The *Magnaporthe oryzae* effector AVR-Pik has evolved through an arms race with the rice immune receptor Pik.

The ancient allele of AVR-Pik effector, AVR-PikD, is recognized by the Pikp immune receptor allele (blue). AVR-PikD likely evaded detection by Pikp through the introduction of a nonsynonymous nucleotide polymorphism that led to emergence of the AVR-PikE allele. The selection pressure imposed by Pik* (purple) and Pikm (yellow) alleles led to further nonsynonymous substitutions and the emergence of AVR-PikA and AVR-PikC alleles. The nucleotide polymorphisms driven by the recognition specificity of the rice immune receptors are shown in bold. Figure adapted from: Białas et al., 2018.

1.8 Aims of this thesis

The main objective of this thesis was to reconstruct the evolutionary history of the Pik-1 and Pik-2 immune receptors of rice and test hypotheses about adaptive evolution of the Pik-1–integrated HMA domain. This work utilises a rich genetic diversity of the *Pik-1*/*Pik-2* genes in grasses as an experimental system for studying the long-term evolution of these NLRs. Phylogenetic and computational analyses revealed that Pik-1 and Pik-2 likely derived from a single ancestral gene pair that emerged before the radiation of the BOP and the PACMAD clades (**Chapter 3**). Furthermore, by looking at the presence/absence polymorphisms of the HMA domain within the Pik-1 homologues, I was able to determine that the HMA integration predates speciation of Oryzinae and has since been under strong diversifying selection. Next, I aimed to understand whether historical emergence of the HMA integration and its diversification within the Pik-1 receptor are a direct adaptation to the rice blast effector AVR-PikD, and to what extent the evolutionary trajectories of Pik-HMAs have been driven by this effector. To challenge these questions, I performed ancestral sequence reconstruction combined with resurrection and functional characterisation of the extinct HMA. Using both in planta and in vitro experiments, I discovered that different allelic variants of Pik-1—Pikp-1 and

Pikm-1—convergently evolved to recognise AVR-PikD (**Chapters 4** and **5**). Using biophysical approaches, I functionally characterised two regions in the Pik-1 HMA that independently evolved towards AVR-PikD binding from the weaker ancestral state. In both cases the HMA domain only recently acquired the capacity to bind the AVR-PikD effector with high affinity, indicating that for most of its evolutionary history the HMA was not subject to selecti pressure imposed by this blast effector. In addition, although Pikp-1 and Pikm-1 receptors evolved to produce a similar phenotypic outcome, they underwent different evolutionary trajectories to do so. This work gives new insights into our understanding of the Pik immune system and provides an example of a robust evolutionary framework for studying mechanisms of protein adaptation.

Chapter 2: Materials and Methods

2.1 Phylogenetic analyses

2.1.1 Identification and phylogenetic analysis of CC-NLRs from grasses

NLR-parser (Steuernagel et al., 2015) was used to identify the NLR sequences from the predicted protein databases of eight representative grass species, *Brachypodium distachyon*, *Oryza brachyantha*, *Oryza sativa*, *Sorghum bicolor*, *Triticum aestivum*, *Zea mays* (downloaded from Ensembl Plants collection), and *Hordeum vulgare* and *Setaria italica* (downloaded from Phytozome v12.1 collection), listed in **Table 2.1**. NLR sequences that were longer than 750 amino acid were screened for features of the NB-ARC and LRR domains, defined by PF00931, PF00560, PF07725, PF13306, and PF13855 pfam models, using HMMER 3.2b2 (Eddy, 1998); signatures of the coiled-coil domain were identified using 'motif16' and 'motif17' defined in NLR-parser. Protein sequences of NLRs that contained at least two of the above features were aligned using MUSCLE v2.8.31 (Edgar, 2004). The proteins comprising of fewer than 60 amino acids N- and C-terminally of the NB-ARC domain, relative to the NB-ARC domain of Pikp-2 (Maqbool et al., 2015), were removed, as were sequences with less than 50% coverage across the alignment. The dataset was further filtered so that for each gene there was only one representative protein isoform —with the exception of sequences from *Brachypodium distachyon* and *Sorghum bicolor* that didn't carry gene identifiers. Filtering resulted in a final list of 3,062 CC-NLRs that were amended with 35 known and functionally characterized NLR-type resistance proteins from grasses, added for the reference (**Table 2.2**).

The amino acid sequences corresponding to the NB-ARC domain of the identified NLRs were aligned using MUSCLE v2.8.31 (Edgar, 2004). The alignment positions with more than 30% data missing were removed from the alignment using QKphylogeny (Moscou, 2019). This revealed a final alignment of 241-amino acids, which was used for a phylogenetic analysis. A maximum likelihood (ML) phylogenetic tree was calculated using RAxML v8.2.11 (Stamatakis, 2014) with bootstrap values (Felsenstein, 1985) based on 1000 iterations and best-scoring JTT likelihood model (Jones et al., 1992) selected by automatic protein model assignment using the ML criterion. Best ML tree was mid-point rooted and visualized using Interactive Tree of Life (iToL) tool v5.5.1 (Letunic and Bork,

2007). The relationships of 28 and 38 proteins that grouped with rice Pikp-1 and Pikp-2, respectively, were further validated. Genetic loci and gene coordinates for each of those NLRs were inspected and, if required, manually reannotated; identifiers of manually reannotated genes were amended with '.n' suffix. For each gene, one splice version was selected and aligned using MUSCLE v2.8.31 (Edgar, 2004). The maximum likelihood phylogenetic trees of Pik-1–related and Pik-2–related NLRs were calculated based on positions within the NB-ARC domain, for which more than 70% of data were present—957 and 1218 nucleotides for Pik-1 and Pik-2, respectively. The trees were generated using RAxML v8.2.11 (Stamatakis, 2014) with bootstrap values (Felsenstein, 1985) based on 1000 iterations and GTRGAMMA substitution model (Tavaré, 1986). Best ML trees were manually rooted based on previously observed relationships and visualized using Interactive Tree of Life (iToL) tool v5.5.1 (Letunic and Bork, 2007).

Table 2.1 List of plant datasets

Species	Cultivar	Version	Assembly accession number	Source
Brachypodium distachyon	Bd21	v3.0	GCA_000005505.4	Ensembl Plants
Dactylis glomerats	Dgl	v1	GCA_007115705.1	NCBI
Hordeum vulgare	Morex	v2	GCA_901482405.1	Phytozome v12.1
Leersia perrieri	*	v1.4	GCA_000325765.3	Ensembl Plants
Oryza barthii	B88	v1	GCA_003020155.1	Ensembl Plants
Oryza brachyantha	IRGC101232	v1.4b	GCA_000231095.2	Ensembl Plants
Oryza glaberrima	*	v1	GCA_000147395.2	Ensembl Plants
Oryza glumaepatula	*	v1.5	GCA_000576495.1	Ensembl Plants
Oryza longistaminata	*	v1.0	GCA_000789195.1	Ensembl Plants
Oryza nivara	*	v1.0	GCA_000576065.1	Ensembl Plants
Oryza punctata	*	v1.2	GCA_000573905.1	Ensembl Plants
Oryza rufipogon	OR_W1943	v1	GCA_000817225.1	Ensembl Plants
Oryza sativa	Nipponbare	IRGSP-1.0	GCA_001433935.1	Ensembl Plants
Setaria italica	Yugu1	v2.2	AGNK01000000.1	Phytozome v12.1
Sorghum bicolor	BTx623	v3	GCA_000003195.3	Ensembl Plants
Triticum aestivum	IWGSC	v1.0	GCA_900519105.1	Ensembl Plants
Zea mays	B73	v4	GCA_000005005.6	Ensembl Plants
Zizania latifolia	*	v1	GCA_000418225.1	NCBI

*information not provided

Table 2.2 List of known and functionally characterized NLR-type resistance proteins from grasses used as reference sequences

Gene	Accession number	Species	Reference
MLA10	AY266445.1	*Hordeum vulgare*	Halterman and Wise, 2004
RGA1-A	KT725812.1	*Secale cereale*	Mago et al., 2015
Os11gRGA5	AB604627.1	*Oryza sativa*	Okuyama et al., 2011
Os11gRGA4	AB604622.1	*Oryza sativa*	Okuyama et al., 2011
Piz-t	DQ352040.1	*Oryza sativa*	Zhou et al., 2006
Pi-ta	AF207842.1	*Oryza sativa*	Bryan et al., 2000
Rpg5	EU883792.1	*Hordeum vulgare*	Brueggeman et al., 2008
LR10	AY270157.1	*Triticum aestivum*	Feuillet et al., 2003
Yr10	AF149112.1	*Triticum aestivum*	Liu et al., 2014
Pib	AB013448.1	*Oryza sativa*	Wang et al., 1999
Pi9	DQ285630.1	*Oryza sativa*	Qu et al., 2006
Rp1-D	XM_008664205.2	*Zea mays*	Collins et al., 1999
Xa1	AB002266.1	*Oryza sativa*	Yoshimura et al., 1998
Pm8	KF572030.1	*Triticum aestivum*	Hurni et al., 2013
Pm3	GU230859.1	*Triticum aestivum*	Bhullar et al., 2010
Rdg2-a	HM124452.1	*Hordeum vulgare*	Bulgarelli et al., 2010
Lr21	FJ876280.1	*Triticum aestivum*	Huang et al., 2009
Pit	AB379815.1	*Oryza sativa*	Hayashi and Yoshida, 2009
Pi5-1	EU869185.1	*Oryza sativa*	Lee et al., 2009
Pi5-2	EU869186.1	*Oryza sativa*	Lee et al., 2009
Pid3	KX791058.1	*Oryza sativa*	Shang et al., 2009
Sr45	LN883757.1	*Triticum aestivum*	Steuernagel et al., 2016
Sr22	LN883743.1	*Triticum aestivum*	Steuernagel et al., 2016
Lr22a	KY064064.1	*Triticum aestivum*	Thind et al., 2017
Pik-1	HM048900_1	*Oryza sativa*	Zhai et al., 2011
Pik-2	ADZ48538.1	*Oryza sativa*	Zhai et al., 2011
Pikh-1	HQ662330_1	*Oryza sativa*	Costanzo and Jia, 2010b
Pikh-2	AET36550.1	*Oryza sativa*	Costanzo and Jia, 2010b
Pikm-1	AB462324_1	*Oryza sativa*	Ashikawa et al., 2008a
Pikm-2	BAG72135.1	*Oryza sativa*	Ashikawa et al., 2008a
Piks-1	HQ662329_1	*Oryza sativa*	Jia et al., 2009
Piks-2	AET36548.1	*Oryza sativa*	Jia et al., 2009
Pikp-1	HM035360.1	*Oryza sativa*	Yuan et al., 2011
Pikp-2	ADV58351.1	*Oryza sativa*	Yuan et al., 2011

2.1.2 Identification and phylogenetic analysis of Pik-1 and Pik-2 homologues

Coding sequences of representative *Pik-1* and *Pik-2* genes were used to identify Pik homologues from cDNA databases of *Oryza barthii*, *Oryza longistaminata*, *Oryza punctata*,

Oryza glumeapatula, *Oryza glaberrima*, *Oryza rufipogon*, *Oryza nivara*, *Leersia pererrii*, *Zizania latifolia*, and *Dactylis glomerata*, listed in **Table 2.1**, using BLAST v2.3.0 (Altschul et al., 1990). For each sequence with BLASTN E-value cutoff <0.01, genetic loci and gene coordinates were inspected and, if necessary, manually reannotated; identifiers of manually reannotated genes were amended with '.n' suffix. Because the *Pik-1* and *Pik-2* genes are known to be genetically linked, each *Pik* locus was further examined for signatures of unpredicted *Pik* gene candidates. Next, coding sequences of the Pik-1 and Pik-2 candidate homologues were aligned using MUSCLE v2.8.31 (Edgar, 2004). Poorly aligned sequences were manually removed from the alignment and excluded from further analysis. The phylogenetic trees were calculated based on positions within the NB-ARC domain, for which more than 70% of data was present—927 and 1239 nucleotides of 46 Pik-1 and 54 Pik-2 candidates, respectively. Maximum likelihood phylogenetic trees were calculated using RAxML v8.2.11 (Stamatakis, 2014) with bootstrap values based on 1000 iterations (Felsenstein, 1985) and GTRGAMMA substitution model (Tavaré, 1986). Best ML trees were manually rooted according to previously observed relationship and visualized using Interactive Tree of Life (iToL) tool v5.5.1 (Letunic and Bork, 2007).

2.2 Molecular evolution methods

2.2.1 Phylogenetic analyses of rice HMA domains and ancestral sequence reconstruction

Selected non-integrated HMA sequences from *Oryza sativa* and *Oryza brachyantha* were obtained from our collaborators at the Iwate Biotechnology Research Center (Kitakami, Japan; unpublished data) and by BLASTP search (Altschul et al., 1990) using Pikp-1 HMA (Pikp-HMA) as a query, respectively. Amino acid and nucleotide alignments were generated using MUSCLE (Edgar, 2004). Neighbour joining (NJ) clustering method (Saitou and Nei, 1987) was used for constructing protein-based or codon-based trees based on JTT (Jones et al., 1992) or Maximum Composite Likelihood substitution models (Tamura and Kumar, 2002), respectively, using 1000 bootstrap test (Felsenstein, 1985), as implemented in MEGA X (Kumar et al., 2018). Maximum likelihood trees were calculated using JTT (Jones et al., 1992) or GTR (Tavaré, 1986) substitution models as implemented in MEGA X software.

Three independent protein sequence alignments, generated with MUSCLE (Edgar, 2004), were used for ancestral sequence reconstruction (**Table 2.3**). Joint and marginal ancestral sequence reconstructions were performed with FastML software (Ashkenazy et al., 2012) using JTT substitution model (Jones et al., 1992), gamma distribution, and 90% probability cut-off to prefer ancestral indel over a character. The reconstruction was performed based on neighbour joining trees (Saitou and Nei, 1987) built with 100 iteration bootstrap method (Felsenstein, 1985). Sequences after marginal reconstruction including indels were used for further analyses.

Table 2.3 HMA sequences used for building phylogenetic trees and ancestral sequence reconstruction (ASR)

Description on the tree	Species	Accession number	Pik-integrated	used for ASR		
				I	II	III
O.barthii_W0042	*O. barthii*	PCR from NIG accession no. W0042	y	n	n	y
O.barthii_W1643	*O. barthii*	PCR from NIG accession no. W1643	y	n	n	y
O.punctata_W1408	*O. punctata*	PCR from NIG accession no. W1408	y	n	n	y
O.barthii_W0698	*O. barthii*	PCR from NIG accession no. W0698	y	n	n	y
O.granulata_W0067B	*O. granulata*	PCR from NIG accession no. W0067(B)	y	n	n	y
O.longistaminata_ W0643	*O. longistaminata*	PCR from NIG accession no. W0643	y	n	n	y
O.officinalis_W0614	*O. officinalis*	PCR from NIG accession no. W0614	y	n	n	y
O.punctata_W1514	*O. punctata*	PCR from NIG accession no. W1514	y	n	n	y
O.rufipogon_W2003	*O. rufipogon*	PCR from NIG accession no. W2003	y	n	n	y
O.minuta_W1328	*O. minuta*	PCR from NIG accession no. W1328	y	n	n	y
LOC102699268	*O. brachyantha*	LOC102699268	y	y	y	y
OBART11G23150	*O. barthii*	OBART11G23150	y	n	n	y
Olongi_KN541092.1	*O. longistaminata*	KN541092.1	y	n	n	y
OPUNC11G19550	*O. punctata*	OPUNC11G19550	y	n	n	y
***Os*Pikp-1**	*O. sativa*	HM035360.1	y	y	y	y
***Os*Pik-1**	*O. sativa*	HM048900_1	y	y	y	y
***Os*Pikh-1**	*O. sativa*	HQ662330_1	y	y	y	y

Table 2.3 HMA sequences used for building phylogenetic trees and ancestral sequence reconstruction (ASR) (continued)

Description on the tree	Species	Accession number	Pik-integrated	used for ASR		
				I	II	III
***Os*Piks-1**	*O. sativa*	HQ662329_1	y	y	y	y
OsPikm-1	*O. sativa*	AB462324.1	y	y	y	y
Ob_LOC102708959	*O. brachyantha*	LOC102708959	n	n	y	y
Ob_LOC102709146	*O. brachyantha*	LOC102709146	n	n	y	y
Ob_LOC102714171	*O. brachyantha*	LOC102714171	n	n	y	y
Ob_LOC102716957	*O. brachyantha*	LOC102716957	n	n	y	y
Ob_LOC102717220	*O. brachyantha*	LOC102717220	n	n	y	y
Os_LOC_Os04g39360	*O. sativa*	LOC_Os04g39360	n	y	y	y
Os_LOC_Os04g39370	*O. sativa*	LOC_Os04g39370	n	y	y	y
Os04g0469000_01	*O. sativa*	Os04g0469000_01	n	y	y	y
Os02g0585200	*O. sativa*	Os02g0585200	n	y	y	y
Os02g0584800_01	*O. sativa*	Os02g0584800_01	n	y	y	y
Os02g0584700_01	*O. sativa*	Os02g0584700_01	n	y	y	y
Os04g0469300_01	*O. sativa*	Os04g0469300_01	n	y	y	y
Os02g0585100	*O. sativa*	Os02g0585100	n	y	y	y
Os02g0584600	*O. sativa*	Os02g0584600	n	y	y	y
OSJNBa0060P14.7_01	*O. sativa*	OSJNBa0060P14.7_01	n	y	y	y
Os04g0464100_01	*O. sativa*	Os04g0464100_01	n	y	y	y
Os02g0582600	*O. sativa*	Os02g0582600	n	y	y	y

I, II, III – different ancestral sequence predictions; y: yes; n: no.

2.2.2 Testing for selection

The rates of synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions per site in pairwise comparisons of protein-coding DNA sequences were estimated using the Yang and Nielsen (2000) method under realistic evolutionary models, as implemented in the YN00 program in the PAML v4.9j package (Yang, 1997). The coding sequence alignments, which were used for the analysis, were generated using MUSCLE v2.8.31 (Edgar, 2004); unless stated otherwise, only positions that showed over 70% coverage across the alignment were used for the analyses.

For selection across the sites of the HMA domain, site models were implemented using the CODEML program in the PAML v4.9j software package (Yang, 1997). The three null models, M0 (one-ratio), M1 (nearly neutral), M7 (beta), and three alternative models, M3 (selection), M2 (discrete), M8 (beta & ω), were tested as recommended by

Yang et al. (2000), and their likelihoods were calculated with the likelihood ratio test. Twice the difference in log likelihood ratio between a null model and an alternative model was compared with the chi-squared (χ^2) distribution; the degrees of freedom were calculated from the difference in the numbers of parameters estimated from the model pairs. The naïve empirical Bayes (NEB) (Yang, 2000; Yang and Nielsen, 1998) or the Bayes empirical Bayes (BEB) (Yang, 2005) were used to infer the posterior probabilities for site classes and to identify amino acids under positive selection. Raw data were extracted and visualized using the *ggplot2* R v3.6.3 package (Ginestet, 2011). Maximum likelihood phylogenetic tree used for the analysis was built with bootstrap values (Felsenstein, 1985) from 1000 iterations using MEGA X software (Kumar et al., 2018), based on coding-sequence alignment, generated with MUSCLE v2.8.31 (Edgar, 2004).

2.3 Molecular biology methods

2.3.1 Golden Gate cloning

Golden Gate assembly was carried out using a protocol modified from literature (Weber et al., 2011). The level 0 restriction-ligation reaction was set up by mixing the following: 100 ng of each insert and level 0 acceptor; 2 U (unit) of *BpiI* (Thermo Fisher Scientific); 4 U T4 DNA ligase (Invitrogen); 1×BSA (NEB) in T4 DNA ligase buffer (Invitrogen), to a final volume of 20 μL. The level 1 restriction-ligation reaction consisted of a mix of: 100 ng of each level 0 module and binary vector; 2 U of *BsaI*-HF (NEB); 4 U T4 DNA ligase (Invitrogen); 1×BSA (NEB) in T4 DNA ligase buffer (Invitrogen), to a final volume of 20 μL. The reaction was incubated for 30 seconds at 37°C, followed by 26 cycles of 5 minutes at 37°C, 5 minutes at 20°C and 10 minutes at 50°C, and finally 10 minutes at 80°C. Afterwards, the reaction was transformed into subclocloning efficiency DH5α chemically competent cells (Invitrogen).

2.3.2 In-Fusion cloning

In-Fusion reactions (Clontech) were performed by mixing 0.5 μL of 5×In-Fusion HD enzyme mix (Clontech), 100 ng of linearized vector, 10 ng of insert, and dH_20

to a total volume of 5 μL, followed by incubation at 50°C for 15 minutes. Next, the reactions were transformed into Stellar chemically competent cells (Clontech).

2.3.3 Traditional cloning

Traditional cloning was performed using CutSmart® restriction enzymes from NEB (New England Biolab) and T4 DNA ligase supplied by Invitrogen, following the manufacturer's instructions. The digested fragments were purified using Zymoclean Gel DNA Recovery Kit (Zymo Research). The ligation reaction was conducted at room temperature overnight. One Shot TOP10 chemically competent *E. coli* cells (Invitrogen) were then transformed with the reaction.

2.3.4 Cloning from genomic DNA

Cloning from genomic DNA was performed by PCR using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) following the manufacturers' instructions. Fragments of the expected size were purified for an agarose gel and cloned into Zero Blunt® TOPO® plasmid (Thermo Fisher Scientific).

2.3.5 Bacterial transformation

Transformations of *E. coli* One Shot TOP10 (Invitrogen), subcloning efficiency DH5α chemically competent cells (Invitrogen), or Stellar chemically competent cells (Clontech) were conducted following the manufacturers' instructions. In brief, competent cells were mixed with ligation products and incubated on ice for up to 30 minutes. Cells were subjected to heat shock by incubation at 42°C for 30 seconds, immediately followed by incubation on ice for 2 minutes. Afterwards, 250 μL of SOC medium (Invitrogen) was added to the cells and the mixture was incubated, with constant agitation, at 37°C for 40 minutes. The cells were transferred on agar plates with lysogeny broth (LB) medium with the appropriate antibiotics (kanamycin 50 μg/mL, spectinomycin 50 μg/mL, or carbenicillin 100 μg/mL) and incubated overnight at 37°C.

Electrocompetent *Agrobacterium tumefaciens* cells, strain GV3101::pMP90, were transformed using an electroporation cuvette with a width of 1 mm and an electroporator

(Biorad) using the following settings: voltage = 1.8 kV, resistance = 200 Ω, capacitance = 25 μF. Immediately after electroporation, 500 μL of LB medium was added to the cells, which were then incubated at 28°C for 40 minutes, with constant agitation. The cells were plated on LB agar plates with the appropriate antibiotics (kanamycin 50 μg/mL and rifampicin 100 μg/mL; spectinomycin 50 μg/mL and rifampicin 100 μg/mL; or carbenicillin 100 μg/mL and rifampicin 100 μg/mL) and incubated at 28°C for approximately 48 hours.

2.3.6 PCR product purification, colony PCR, and plasmid preparation

PCR products were purified using a Zymoclean Gel DNA Recovery Kit (Zymo Research). Colony PCR was performed using DreamTaq DNA polymerase following the manufacturer's instructions (Thermo Fisher Scientific). Plasmid extraction was carried out using QIAprep Spin Miniprep Kit according to the manufacturer's instructions (Qiagen).

2.4 Plant material

Wild type *Nicotiana benthamiana* plants were grown in a controlled growth chamber with 22–25°C, 45–65% humidity, and 16/8-hour light/dark cycle.

2.5 Cloning

2.5.1 Identification and cloning of *Pik-1* and *Pik-2* from *Oryza brachyantha*

Genomic DNA materials of 16 *Oryza brachyantha* accessions were ordered from Wild Rice Collection 'Oryzabase' (**Table 2.4**) (Kurata and Yamazaki, 2006). The accessions were first screened for deletion within the *Pik-2* gene, present in a reference genome of *Oryza brachyantha* (Chen et al., 2013). Selected accessions were used to amplify full-length *Pik-1* and *Pik-2* genes using primers designed based on *Oryza brachyantha* genome sequence (**Table 2.5**). The PCRs were run on 2% agarose gels to check amplification and product size against positive controls. Fragments of the expected size were further gel purified, cloned into Zero Blunt® TOPO® plasmid (Thermo Fisher Scientific), and sequenced.

Table 2.4 List of *Oryza brachyantha* accessions

Accession	Country of origin	Used for *Pik-2* cloning	Used for *Pik-1* cloning
W0654	Sierra Leone	Full-length	Full-length
W0655	Sierra Leone	Not sequenced	Full-length
W0656	Guinea	Fragment	Not amplified
W1057	Guinea	Fragment	Not amplified
W1401	Sierra Leone	Fragment	Not amplified
W1402	Sierra Leone	Fragment	Not amplified
W1403	Sierra Leone	Not sequenced	Not amplified
W1404	Sierra Leone	Full-length	Full-length
W1405	Sierra Leone	Full-length	Full-length
W1407(B)	Mali	Full-length	Full-length
W1703	Mali	Full-length	Full-length
W1705	Mali	Full-length	Full-length
W1706	Chad	Fragment	Not amplified
W1708	Cameroun	Fragment	Not amplified
W1711	Cameroun	Fragment	Not amplified
W1712	Cameroun	Fragment	Not amplified

Coding sequence of *Ob*Pik-1 (accession number: LOC102699268) was synthesised as a level 0 module for Golden Gate cloning (Weber et al., 2011) by GENEWIZ. Using this module, expression construct featuring an N-terminal tag was generated by Golden Gate assembly with pICSL12008 (35S + Ω promoter, TSL SynBio), pICSL30007 (N-terminal 6×HA, TSLSynBio), and pICH41414 (35S terminator, TSL SynBio), into the binary vector pICH47732 (Addgene no. 48001). All constructs were verified by DNA sequencing. Sequences of primers used for *Ob*Pik-1 and *Ob*Pik-2 identification and cloning are listed in **Table 2.4**.

2.5.2 Identification and cloning of the Pik-1–integrated HMA domains from wild rice relatives

Genomic DNA materials of one to three accessions of 18 wild rice species—*Oryza australiensis*, *Oryza barthii*, *Oryza brachyantha*, *Oryza eichingeri*, *Oryza glumaepatula*, *Oryza grandiglumis*, *Oryza granulata*, *Oryza latifolia*, *Oryza longiglumis*, *Oryza longistaminata*, *Oryza meridionalis*, *Oryza meyeriana*, *Oryza minuta*, *Oryza officinalis*, *Oryza punctata*, *Oryza rhizomatis, Oryza ridleyi*, *Oryza rufipogon*—were ordered from Wild Rice Collection 'Oryzabase' (Kurata and Yamazaki, 2006) and used for amplification of the Pik-1 integrated HMA (**Table 2.6**).

The forward 5'-AGGGAGCAATGATGCTTCACGA-3', and reverse 3'-TTCTCTGGCAACCGTTGTTTTGC-5', primers were designed using the alignment of the *Os*Pikp-1 and OBRAC11G13570.1 sequences and used in PCR. The amplicons were run on 2% agarose gels to check amplification and product sizes against positive controls. Fragments of 450–720 bp in size were gel-purified, cloned into Zero Blunt® TOPO® plasmid (Thermo Fisher Scientific), and sequenced using M13 universal primers (Thermo Fisher Scientific). Genotyping was performed twice and only sequences that did not show ambiguity between sequencing runs were selected for further analyses.

Table 2.5 List of primers used for Pik-1 and Pik-2 cloning

Primer name	Sequence (5′-3′)	Usage in this study	Reference
ObPik2_del_F	*TGAAGCAGATCCGAGACATAGCCT*	Screening for deletion	This study
ObPik2_del_R	*TACCCTGCTCCTGATTGCTGACT*	Screening for deletion	This study
ObPik-2_gene_F1	*GTCCACGCTTGATCTGAACTGT*	Pik-2 cloning	This study
ObPik-2_gene_R2	*TGTGGCACCATAATTATAGCAATCCC*	Pik-2 cloning	This study
ObPik-1_gene_F1	*GGCACCTGGCACTAACTAGGTA*	Pik-1 cloning	This study
ObPik-1_gene_R	*TCAACGCAAAAGGGGATCCGA*	Pik-1 cloning	This study
ObPik-1_gene_F2	*GATGCTCCTTGGCTTCTTTCTTG*	Pik-1 sequencing	This study
ObPik-1_gene_R1	*GAGTTCAAGTTCTACGTCGGCC*	Pik-1 sequencing	This study
ObPik-1_gene_F3	*GTTTTCCTCACGTCCAGAGTCTG*	Pik-1 sequencing	This study
ObPik-1_gene_F4	*CTCATTATCTTCTAGGTCCAACACTCG*	Pik-1 sequencing	This study
ObPik-1_gene_F5	*CAGACGACCCCTTTCGATTCTG*	Pik-1 sequencing	This study
ObPik-1_gene_F6	*GGAATGGACTTTCTGATGACTTCCC*	Pik-1 sequencing	This study
ObPik-1_gene_F7	*ATATACATATACTGCAAGCCTTTGCAA G*	Pik-1 sequencing	This study
ObPik-1_gene_F8	*GTCATATAATTCTTTGGCAACCGTTGT C*	Pik-1 sequencing	This study
ObPik-1_gene_F9	*TGTCAACGATCCCTCCATGTT*	Pik-1 sequencing	This study
ObPik-1_gene_R	*TCAACGCAAAAGGGGATCCGA*	Pik-1 sequencing	This study
ObPik-2_gene_F2	*CCATTGCCACAGCATTGTACAG*	Pik-2 sequencing	This study
ObPik-2_gene_F3	*TCCTCTTGATCGACGATATTTGGTC*	Pik-2 sequencing	This study
ObPik-2_gene_F4	*GATACTGGATTGTTGTTACAATGATTT G*	Pik-2 sequencing	This study
ObPik-2_gene_F5	*CACCAAGCAATAAAGTCCGTCG*	Pik-2 sequencing	This study
ObPik-2_gene_F6	*CAAGAAAAAGGTAAAGGCGCAATGAA*	Pik-2 sequencing	This study

*size depends on accession

Table 2.6 List of wild rice accessions

Accession number	Species	Genome	Country of origin
W0008	*Oryza australiensis*	EE	Australia (SE Canberra)
W1628	*Oryza australiensis*	EE	Australia (N)
W1643	*Oryza barthii*	AA	Botswana
W1605	*Oryza barthii*	AA	Nigeria
W0042	*Oryza barthii*	AA	NA
W0698	*Oryza barthii*	AA	Guinea
W0654	*Oryza brachyantha*	FF	Sierra Leone
W1526	*Oryza eichingeri*	CC	Uganda
W1171	*Oryza glumaepatula*	AA	Cuba
W2203	*Oryza glumaepatula*	AA	Brasil (S)
W1480(B)	*Oryza grandiglumis*	CCDD	Brasil (N)
W0005	*Oryza granulata*	GG	Sri Lanka
W0067(B)	*Oryza granulata*	GG	Thailand
W0542	*Oryza latifolia*	CCDD	Mexico
W1539	*Oryza latifolia*	CCDD	Argentina (N)
W1228	*Oryza longiglumis*	HHJJ	Indonesia
W1504	*Oryza longistaminata*	AA	Tanzania
W1540	*Oryza longistaminata*	AA	Republic of Congo
W0643	*Oryza longistaminata*	AA	The Gambia
W2081	*Oryza meridionalis*	AA	Australia (N)
W2112	*Oryza meridionalis*	AA	Australia (NE)
W1354	*Oryza meyeriana*	GG	Malaysia
W1328	*Oryza minuta*	BBCC	Philippines
W0614	*Oryza officinalis*	CC	Myanmar
W1200	*Oryza officinalis*	CC	Philippines
W1408	*Oryza punctata*	BBCC	Nigeria
W1514	*Oryza punctata*	BBCC	Kenya
W1808	*Oryza rhizomatis*	CC	Sri Lanka
W0001	*Oryza ridleyi*	HHJJ	Thailand
W2035	*Oryza ridleyi*	HHJJ	Philippines
W2003	*Oryza rufipogon*	AA	India (SW)
W1715	*Oryza rufipogon*	AA	Chin (Beijing)

2.5.3 Cloning of AVR-PikD

AVR-PikD, previously cloned by Maqbool et al. (2015), was amplified from pDONR221 plasmid using primers listed in **Table 2.7**; the reverse primer was used to introduce C-terminal FLAG tag. The PCR product was purified and digested with *PacI* and *NotI* restriction enzymes, according to manufacturer's instructions (NEB). Following

another round of purification, the PCR product was ligated with linearized pTRBO destination vector (Lindbo, 2007), generating AVR-PikD:FLAG expression construct; the construct was verified by DNA sequencing.

Table 2.7 List of primers used for AVR-PikD cloning

Primer name	Sequence (5′–3′)	Reference
AVRPikD_PacI_F	*GGTTAATTAACCATGGAAACGGGTAATAAATACATTGAAAA ACGTG*	This study
AVRPikD_FLAG_ NotI_R	*AAGCGGCCGCTCATTTATCGTCATCGTCCTTATAATCAGC TGCTGCAGCGTCGAGGTCATGGTCCTTATAGTCTCCGTCA TGGTCCTTATAATCGAAGCCCGGACGTTTTTTACCC*	This study

2.5.4 Cloning of *Os*Pikp-1

Sequences of primers used for Pik-1 cloning are listed in **Table 2.8**. The rice Pikp-1, previously cloned by Maqbool et al. (2015), was amplified from pCambia1300:AscI plasmid and domesticated to remove internal *BsaI* and *BpiI* restriction enzyme recognition sites. Domestication was performed using site-directed mutagenesis by inverse PCR with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). The amplicons were purified and assembled using Golden Gate method (Weber et al., 2011) in the level 0 pICH41308 destination vector (The Sainsbury Laboratory [TSL] SynBio) for follow-up Golden Gate cloning. The N-terminally tagged HA:Pikp-1 expression construct was generated by Golden Gate assembly with pICSL12008 (35S + Ω promoter, TSL SynBio), pICSL30007 (N-terminal 6×HA, TSL SynBio), and pICH41414 (35S terminator, TSL SynBio) modules, into the binary vector pICH47732 (Addgene no. 48001). Using the same set of Golden Gate modules, Pikp-1$_{E230R}$ mutant, which was generated in the Banfield lab at the John Innes Centre (Norwich, UK), was subcloned into the same binary vector, generating N-terminally tagged HA:Pikp-1$_{E230R}$ expression construct. All constructs were verified by DNA sequencing.

2.5.5 Cloning of *Os*Pik-1:ancHMA fusions

In order to generate Pikp-1:ancHMA fusions, individual domains of Pikp-1 were amplified and cloned into pCR8/GW/D4TOPO (Invitrogen) plasmid as level 0 modules. Such modules were used in Golden Gate assembly (Weber et al., 2011) with ancHMA

variants, pICSL12008 (35S + Ω promoter, TSL SynBio), pICSL30007 (N-terminal 6×HA, TSLSynBio), pICH41414 (35S terminator, TSL SynBio), and pICH47732 binary vector (Addgene no. 48001), generating Pikp-1:ancHMA expression constructs featuring N-terminal HA tag. The ancestral HMA variants—corresponding to 186–260 residues of the full-length Pikp-1—had been synthesised as level 0 modules for Golden Gate cloning by GENEWIZ (South Plainfield, NJ, USA).

To enable faster cloning of subsequent Pikp-1:ancHMA fusions, two Golden Gate level 0 acceptor plasmids, p41308-PikpN and p41308-PikpC, were generated by The Sainsbury Laboratory SynBio team (Norwich, UK); the acceptors allowed HMA insertion in a single Golden Gate level 0 reaction, generating full-length Pikp-1 construct with or without a stop codon, respectively. The ancestral HMA mutants—ancHMA$_{AMEGNND}$, ancHMA$_{LY}$, ancHMA$_{PI}$, ancHMA$_{LVKIE}$, and the single mutants within the LVKIE region of the ancHMA—were synthetized by GENEWIZ (South Plainfield, NJ, USA) and subcloned into p41308-PikpN and p41308-PikpC plasmids for follow-up cloning. Two of the ancHMA mutants, ancHMA$_{IVQVE}$ and ancHMA$_{LVKIV}$, were generated using site-directed mutagenesis by inverse PCR with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) and cloned into the same acceptor plasmids. Using the p41308-PikpN modules, HA:Pikp-1:ancHMA expression constructs were generated by Golden Gate assembly with pICSL12008 (35S + Ω promoter, TSL SynBio), pICSL30007 (N-terminal 6×HA, TSL SynBio), and pICH41414 (35S terminator, TSL SynBio) into the binary vector pICH47732 (Addgene no. 48001). To generate C-terminally tagged expression constructs, the p41308-PikpC modules were assembled with pICSL13004 (Mas promoter, TSL SynBio), pICSL50001 (C-terminal HF, TSL SynBio), and pICH77901 (Mas terminator, TSL SynBio), by Golden Gate method into the same binary vector. All constructs were verified by DNA sequencing.

To generate Pikm-1:ancHMA fusions, ancHMA N2-I, ancHMA$_{EMVKE}$, ancHMA$_{FFE}$, ancHMA$_{STSN}$, ancHMA$_{VH}$, and ancHMA$_{IVDPM}$ were synthesised by GENEWIZ (South Plainfield, NJ, USA) as Golden Gate modules. The ancHMA$_{EMANK}$ mutant was generated by amplification and fusion of the N-terminus of ancHMA$_{EMVKE}$ construct and the C-terminus of N2-I ancHMA variant. All ancHMA constructs corresponded to 187–264 residues of the full-length Pikm-1 protein and were subsequently assembled with p41308-PikmN (TSL SynBio) or p41308-PikmC (TSL SynBio) level 0 acceptors

to generate Pikm-1:ancHMA fusions with or without a stop codon, respectively. Obtained modules were then used to generate Pikm-1:ancHMA expression constructs, featuring either N-terminal HA of C-terminal HF tags, by Golden Gate assembly using the same set of modules as previously used for Pikp-1 and pICH47732 binary vector; constructs were verified by DNA sequencing. Sequences of primers used for cloning *Os*Pik-1 fusions are listed in **Table 2.8**.

Table 2.8 List of primers used for Golden Gate cloning of Pik-1 constructs

Primer name	Sequence	Usage in this study	Reference
Pikp1_level0_F	*GGTCTCTAATGGAGGCGGCTGCC ATGGCCGTAA*	Golden Gate cloning and domestication	This study
Pikp1_dom_1_R	*GGTCTCTGTCCAAGGTACCTCACT CTGAGGCTTAACTGTTCGCATA*	Golden Gate cloning and domestication	This study
Pikp1_dom_1_F	*GGTCTCTGGACTCAAGGGAACGC GGATCCGAAAG*	Golden Gate cloning and domestication	This study
Pikp1_dom_2_R	*GGTCTCTTCCTCATGGAAGGCTAT CCTTGGCACCC*	Golden Gate cloning and domestication	This study
Pikp1_dom_2_F	*GGTCTCTAGGACGCCATGCCCAA CCTCAAGCTTC*	Golden Gate cloning and domestication	This study
Pikp1_level0_R	*GGTCTCTAAGCCTAGCTAGTAGTT TCTGTTTGAATTTCAATATCTGCTA CTCG*	Golden Gate cloning and domestication	This study
OsCC_lev0_F	*AAGGTCTCTAATGGAGGCGGCTG CCATGG*	Golden Gate cloning Pikp-1:ancHMA	This study
OsCC_GGTA_lev 0_R	*AAGGTCTCTTACCTGTCTTTCTGA GCGGGAAATGC*	Golden Gate cloning Pikp-1:ancHMA	This study
OsNB_CAAG_lev 0_F	*AAGGTCTCTCAAGAATAAAGATGT GAAGGAGACGACGCC*	Golden Gate cloning Pikp-1:ancHMA	This study
OsNB_lev0_R	*AAGGTCTCTAAAGTAACCTTCTGC TTCTTTCTCTTCCGACAC*	Golden Gate cloning Pikp-1:ancHMA	This study
OsLRR_lev0_R	*AAGGTCTCTAAGCCTAGCTAGTAG TTTCTGTTTGAATTTCAATATCTG*	Golden Gate cloning Pikp-1:ancHMA	This study
OsLRR_lev0_F	*AAGGTCTCTCTTTGGCGAGCTTAT TAACAGAGGATGGATTACG*	Golden Gate cloning Pikp-1:ancHMA	This study
mEM-ANK_Nterm_F	*AAGAAGACATCAGAATAATGGGAG GTGAAATG*	Golden Gate cloning $ancHMA_{EMANK}$	This study
mEM-ANK_Nterm_R	*TTGAAGACATAGGTCACCGGCGAT TGCAAC*	Golden Gate cloning $ancHMA_{EMANK}$	This study
mEM-ANK_Cterm_F	*AAGAAGACATACCTCAGGGACCAG GTTGTG*	Golden Gate cloning $ancHMA_{EMANK}$	This study
mEM-ANK_Cterm_R	*TTGAAGACATCATCCTTGTTGGCC TG*	Golden Gate cloning $ancHMA_{EMANK}$	This study
N2_IVQVE_Nter_F	*AAGAAGACATGACCGGTATGAAGC AAAAAATCG*	ancHMA site-directed mutagenesis	This study

Table 2.8 List of primers used for Golden Gate cloning of Pik-1 constructs (continued)

Primer name	Sequence	Usage in this study	Reference
N2_IVQVE_Nter_R	*TTGAAGACATAGGTCACCGACGATTGCAACC*	ancHMA site-directed mutagenesis	This study
N2_IVQVE_Cter_F	*AAGAAGACATACCTCAGGGACCAGGTTGAGGTG*	ancHMA site-directed mutagenesis	This study
N2_IVQVE_Cter_R	*TTGAAGACATCATCTTTGTTGGCCTGGC*	ancHMA site-directed mutagenesis	This study
N2_LVKIV_Nter_R	*TTGAAGACATCTCTTAGATCACCTACGAGTGCAACC*	ancHMA site-directed mutagenesis	This study
N2_LVKIV_Cter_F	*TTGAAGACATAGAGACAAGATAGTGGTGGTCGGTG*	ancHMA site-directed mutagenesis	This study

2.5.6 HMA cloning for in vitro studies

The ancHMA mutants were amplified from Golden Gate level 0 modules using primers listed in **Table 2.9** and cloned into pOPINM vector featuring N-terminal 6×His and MBP tags, using In-Fusion cloning (Berrow et al., 2007); the constructs were verified by DNA sequencing.

2.6 Biochemistry and biophysics methods

2.6.1 In planta transient protein expression

Transient gene expression in planta was conducted by delivering T-DNA constructs within *Agrobacterium tumefaciens* strain GV3101::pMP90 into *Nicotiana benthamiana* leaves as described before (Win et al., 2011). Briefly, overnight cultures of *A. tumefaciens* carrying expression vector of appropriate constructs were mixed in 1:1 ratio and diluted in agroinfiltration buffer (10 mM 2-[N-morpholine]-ethanesulfonic acid [MES], 10 mM $MgCl_2$, and 150 μM acetosyringone, pH 5.6) to a final OD600 of 0.3, unless stated otherwise. Upper leaves of 4–5-weeks-old *N. benthamiana* plants were agroinfiltrated, and the leaf tissue was collected 3 days after infiltration.

Table 2.9 List of primers used for ancHMA cloning for in vitro studies

Primer name	Sequence	Usage in this study	Reference
AB_N2_pOPINM_0_R	*ATGGTCTAGAAAGCTTTAGCTGACC TGCAACAACTCCGCG*	In-Fusion cloning of ancHMA	This study
AB_E230R_pOPINM_0_R	*ATGGTCTAGAAAGCTTTAGCTGACC TGCAGCAACTCCGC*	In-Fusion cloning of ancHMA	This study
AB_Pikp_opt_pOPINM_0_R	*ATGGTCTAGAAAGCTTTACGAGACT TGCAGCAGTTCCGCATC*	In-Fusion cloning of ancHMA	This study
AB_N12_pOPINM_F	*AAGTTCTGTTTCAGGGCCCGGGTAT GAAGCAAAAAATCGTGATCAAGGTT CCC*	In-Fusion cloning of ancHMA	This study
AB_Pikp_opt_pOPINM_+2_F	*ATGGTCTAGAAAGCTTTAGGTCTGA AACAAAAAATCGTTATCAAAGTCGC AATGG*	In-Fusion cloning of ancHMA	This study
AB_N2-for_m_5_pOPIN_F	*AAGTTCTGTTTCAGGGCCCGGGAG GTATGAAGCAAAAAATCGTGATC*	In-Fusion cloning of ancHMA	This study
AB_N2-for_m_5_pOPIN_R	*ATGGTCTAGAAAGCTTTAACCCTTG TTGGCCTGGC*	In-Fusion cloning of ancHMA	This study
AB_N2-EMVKE_pOPIN_F	*AAGTTCTGTTTCAGGGCCCGGGAG GTGAAATGCAAAAAATCGTGATC*	In-Fusion cloning of ancHMA	This study
AB_N2-EMVNK_pOPIN_R	*ATGGTCTAGAAAGCTTTAACCTTCCT TTACCTGGCTGAC*	In-Fusion cloning of ancHMA	This study

2.6.2 Plant total protein extraction

Total protein extraction was carried out as described previously (Win et al., 2011). Briefly, *N. benthamiana* leaf tissue was ground into a powder in liquid nitrogen using a mortar and pestle. Leaf powder was mixed at 3:1 ratio (w/v) with ice-cold extraction buffer GTEN (10% [w/v] glycerol; 25 mM Tris-HCl, pH 7.5; 1 mM EDTA; 300 mM NaCl) supplemented with 2% (w/v) PVPP, 10 mM dithiothreitol (DTT), 1% (v/v) protease inhibitor cocktail (Sigma), and 0.2% (v/v) Nonidet P-40. After centrifugation at 6000×g at 4°C for 20–30 minutes, the supernatant was transferred to a fresh tube and centrifuged again in the same conditions. The supernatant was passed through 0.45 μM filter, resulting in the total protein extract. For SDS-PAGE electrophoresis, the protein extract was mixed with protein loading dye (4×final concentration: 200 mM Tris-HCl, pH 6.8; 2.5% [v/v] glycerol; 0.2% [w/v] bromophenol blue; 4% [w/v] SDS) and incubated at 95°C for 10 minutes before electrophoresis.

2.6.3 Co-immunoprecipitation

Co-immunoprecipitation was performed as previously described (Win et al., 2011) using affinity chromatography with Anti-HA Affinity Matrix (Roche). The immunoprecipitation was performed by adding 30 µL of beads, unless stated otherwise, to 4–5 mL of total protein extract, and mixing the extract on a rotary mixer at 4°C. After 1–1.5-hour incubation the sample was centrifuged at 800×g at 4°C for 1 min and washed as followed: supernatant was discarded, the resin was resuspended in 1 mL of the IP buffer (GTEN supplemented with Nonidet P-40). Washing step was repeated five times. After the last wash, the remaining liquid was aspirated using a needle attached to a 1 mL syringe; the beads were resuspended in 30 µL of 1.5×loading dye with 10 mM DTT and eluted by incubating at 70°C for 10 minutes.

2.6.4 SGS-PAGE electrophoresis

For Western blot analysis, the SDS-PAGE electrophoresis was performed in Tris-glycine buffer (25 mM Tris; 250 mM glycine, pH 8.3; 0.1% [w/v] SDS) for approximately two hours at 120 V using commercial 4–20% SDS-PAGE gels (Bio-Rad). For in vitro analysis, commercial 16% RunBlue™ TEO-Tricin SDS gels (Expedeon) were used for electrophoresis in RunBlue™ SDS Running Buffer (Expedeon) for approximately 40 minutes at 150 V. PageRuler or PageRulerPlus (Fermentas) were used as protein size markers.

2.6.5 Immunoblot analysis

Following, SDS-PAGE electrophoresis, the proteins were transferred onto a polyvinylidene difluoride (PVDF) membrane using a Trans-Blot Turbo transfer system according to the manufacturer's instructions (Bio-Rad). The membrane was blocked with 5% milk in the TBST (Tris-buffered saline with 1% Tween 20) buffer for minimum of 30 minutes and probed with appropriate antisera. HA-probe (F-7) horseradish peroxidase (HRP)-conjugated (Santa Cruz Biotech) was used for a single-step detection of HA tag. FLAG detection was carried using monoclonal ANTI-FLAG® M2 (Sigma) and anti-mouse HRP-conjugated antibodies in a two-step FLAG detection. A two-step detection of Myc was performed using anti-Myc (A-14, Santa Cruz Biotechnology) and

anti-rabbit HRP-conjugated antibodies. Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific) or SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) were used for detection. Membranes were imaged using ImageQuant LAS 4000 luminescent imager (GE Healthcare Life Sciences). Equal loading was checked by staining PVDF membranes with Pierce Reversible Protein Stain Kit (Thermo Fisher Scientific), homemade Ponceau S stain (Ponceau S in 1% [v/v] acetic acid), or homemade Coomassie Blue Staining solution (10% acetic acid and 50% methanol).

2.6.6 Heterologous protein production and purification

For surface plasmon resonance analysis, heterologous protein production and purification were performed as previously described (De la Concepcion et al., 2018). Briefly, pOPINM constructs encoding HMAs were produced in *E. coli* SHuffle cells (Lobstein et al., 2012). The cells were grown as pre-cultures in 100 mL volumes of LB media at 37°C overnight with constant agitation; the pre-cultures were used for inoculation 1 L volumes of auto-induction media (Studier, 2005), which were grown at 30°C for 4–6 hours before induction at 18°C overnight. The cells were harvested by 15-minute centrifugation at 5,500×g and re-suspended in A1 buffer (50 mM Tris-HCl, pH 7.5; 5% [v/v] glycerol; 500 mM NaCl; 50 mM glycine; 20 mM imidazole; and EDTA-free protease inhibitor [Rosche]), and lysed by sonification for 4 minutes in pulses (1 second on, 3 seconds off). The lysate was clarified by centrifugation at 18,000×g for 20 minutes at 4°C and used for protein tandem purification by immobilized metal affinity chromatography and gel filtration (IMAC-GF). The samples were injected onto a Ni^{2+}-NTA column (GE Healthcare) connected to ÄKTA Pure purification system (GE Healthcare). Proteins were step-eluted with elution buffer (A1 supplemented with 500 mM imidazole) and eluted His-tagged proteins were applied to Superdex 75 26/60 gel filtration column (GE Healthcare) pre-equilibrated with A2 buffer (20 mM HEPES, pH 7.5; 150 mM NaCl) augmented with 1 mM Tris(2-carboxyethy)phosphine hydrochloride (TCEP). To remove purification tags, proteins were treated with 3C protease at 10 μg/mg protein and incubated overnight at 4°C; the tag capture was carried out by tandem Ni^{2+}-NTA and MBP-Trap HP columns (GE Healthcare). The flow-through, concentrated as appropriate, was injected onto Superdex 75 26/60 gel

filtration column for final purification and buffer exchange into A2 buffer augmented with 1 mM TCEP. AVR-PikD effector with non-cleavable C-terminal 6×His tag was produced and purified as previously described (Maqbool et al., 2015).

For crystallisation, ancHMA, ancHMA$_{LVKIE}$, and ancHMA$_{EMVKE}$ encoded in pOPINM, as well as AVR-PikD expressed in pOPINE vector, were grown and harvested as described above. Following initial protein purification by IMAC-GF and removal of the solubility tags, the ancHMA–AVR-PikD protein pairs were combined and subsequently treated as single samples for the final gel filtration purification in A2 buffer. The concentration of proteins was measured by protein absorbance at 205 nm or by colorimetric method using Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) according to manufacturer's instructions. The purity of the proteins was accessed by running the samples on the 16% SDS-PAGE gels and staining with InstantBlue™ (Expedeon). Protein intact masses were determined by LC-MS performed by The Sainsbury Laboratory core Proteomics facilities (Norwich, UK).

2.6.7 Surface plasmon resonance

Surface plasmon resonance (SPR) experiments determining binding within the LVKIE and EMVKE regions were performed in SPR buffer 1 (50 mM HEPES, pH 7.5; 300 mM NaCl; and 0.1% Tween 20) and SPR buffer 2 (50 mM HEPES, pH 7.5; 820 mM NaCl; and 0.1% Tween 20), respectively, at 25°C using Biacore T200 (GE Healthare), unless stated otherwise. A multicycle kinetics approach was used to study the interaction between 6×His-tagged AVR-PikD (ligand) immobilised on the Series S Sensor Chip NTA (GE Healthcare) and the HMA constructs (analytes) flowed over the effector at a flow rate of 30 μL/min. For each kinetic cycle, the chip was washed with the SPR buffer and activated with 30 μL of 0.5 mM NiCl prior immobilisation of AVR-PikD. The HMA proteins in different concentrations (ranging from 0 to 600 nM, depending on the construct) were injected over both reference and sample cells for 360 s, and dissociation was recorded for further 120 s, unless stated otherwise. Between each cycle the sensor chip was regenerated with 30 μL of 0.35 M EDTA. To correct for bulk refractive index changes or machine errors, the response for each measurement was subtracted by the response in the reference cell and the response in buffer-only run. Steady state or kinetics-based affinity for HMA binding to immobilised AVR-PikD was determined

from multi-cycle kinetics curves using the Biacore Insight Evaluation Software (GE Healthcare). For the kinetics-based method the association/dissociation constants (K_D) was calculated with a one-to-one model.

Due to insufficient quality of data, it was not possible to determine the K_D for some of the interactions. In order to compare level of binding for those constructs, the side-by-side multicycle kinetics was performed for Pikp– and Pikm–ancHMA constructs at 50 nM, 200 nM, and 400 nM concentrations, using 120 s contact time and 120 s disassociation time. The theoretical maximum responses (R_{max}) normalized for the amount of ligand immobilized on the chip were calculated and the level of binding was expressed as a percentage of R_{max} (%R_{max}). Each experiment was repeated a minimum of two times, with similar results. The data were visualised using *ggplot2* R package (Ginestet, 2011).

2.6.8 Crystallisation and data collection

All crystallisation screens were performed at 18°C using a sitting-drop vapour diffusion technique. Drops composed of 0.3 μL of protein solution and 0.3 μL of reservoir solution were set up in MRC 96-well crystallisation plates (Molecular Dimensions), which were dispensed using an Oryx Nano or an Oryx8 robot (Douglas Instruments). Screenings with microseeding were performed by dispensing 0.2 μL of protein solution, 0.1 μL of seed stock, and 0.3 μL of reservoir solution. Seed stocks were prepared by placing crystals in 50 μL of reservoir solution with a single glass bead and crashing the crystals by mixing for approximately 2 minutes using a vortex mixer. The crystal growth was monitored using The Minstrel Desktop Crystal Imaging System (Rikagu). Grown crystals were transferred into precipitation solution augmented with 25% (v/v) cryoprotectant—ethylene glycol or glycerol—and flash-cooled in liquid nitrogen using LithoLoops (Molecular Dimensions). The X-ray datasets were collected at the Diamond Light Source using beamline DLS-i04 (Didcot, UK).

Crystals of ancHMA$_{LVKIE}$/AVR-PikD were grown using purified protein at concentration of ~20 mg/mL. Three different crystallisation screens were used: PEGs Suite (Qiagen), Morpheus® (Molecular Dimensions), and KISS (in house, **Table A.1.1**). The crystals of the complex grew after 24–48 hours in well E2 (20% [w/v] PEG 3350; 0.2 M ammonium sulphate; sodium acetate, pH 4) of the KISS screen, and in wells G11

(20% [w/v] PEG 3350, 0.2 M ammonium sulphate) and H7 (20% [w/v] PEG 3350, 0.2 M ammonium phosphate) of the PEGs Suite screen. Based on the screen results, optimisation of crystallisation conditions was set up as shown in **Figure A.1.1**. The X-ray datasets for 11 crystals from three precipitation conditions (condition 1: ~15% [w/v] PEG 3350, 0.1 M ammonium sulphate, and acetate [pH 4]; condition 2: ~15% [w/v] PEG 3350, 0.2 M ammonium sulphate; condition 3: ~15% [w/v] PEG 3350, 0.2 M tri-sodium citrate) were collected. Crystals used for structure resolution appeared in 14.41% (w/v) PEG 3350 and 0.2 M tri-sodium citrate, and were frozen using 25% ethylene glycol.

Crystallisation screens of the ancHMA$_{EMVKE}$/AVR-PikD complex were performed with purified protein at concentrations of 21 mg/mL, 19 mg/mL, or 10.5 mg/mL, and the following 96-well plate screens: PEGs Suite (Qiagen), Morpheus® (Molecular Dimensions), KISS (in house, **Table A.1.1**), Structure (Molecular Dimensions), and JCSG-plus (Molecular Dimensions) as well as custom optimisation conditions (**Figure A.1.1**). Four crystals, which were sent for X-ray data collection, were grown using protein at concentration of 21 mg/mL in H5 well (20% [w/v] PEG 3350, 0.2 M potassium phosphate) of the PEGs Suite screen, and frozen in liquid nitrogen with 25% (v/v) ethylene glycol cryoprotectant.

For the ancHMA/AVR-PikD complex two alternative ancHMA constructs were used, ancHMA (corresponding to 186–258 residues of the full-length Pikp-1) or ancHMA+5 (corresponding to 186–263 residues of the full-length Pikp-1). Crystallisation screens were performed using: PEGs Suite (Qiagen), Morpheus® (Molecular Dimensions), KISS (in house, **Table A.1.1**), Structure (Molecular Dimensions), JCSG-plus (Molecular Dimensions), and PGA (Molecular Dimensions) screens, and purified protein of concentration ranging from 7.7 to 18 mg/mL. Crystallisation conditions were further expanded using the most promising precipitants as shown in **Figure A.1.1**. To promote crystal growth, Structure (Molecular Dimensions) and Morpheus (Molecular Dimensions) screens were dispensed with 3.5 mg/mL of protein with addition of seed stock of ancHMA$_{LVKIE}$/AVR-PikD crystals. Irregular crystals were grown in wells A1 (30% [w/v] PEG 3350; 0.1 M sodium acetate, pH 4.0), E3 (30% [w/v] PEG 3350; 0.2 M ammonium sulphate; 0.1 M sodium acetate, pH 4.0), F3 (30% [w/v] PEG 3350; 0.1 M ammonium sulphate; 0.1 M HEPES, pH 7.0) of the KISS screen, and well A4 of the

Morpheus screen composed of: 0.06 M divalents (0.3 M magnesium chloride hexahydrate, 0.3 M calcium chloride dihydrate); 0.1 M buffer system 1 (1 M imidazole; MES monohydrate [acid]), pH 6.5; 50% (v/v) precipitant mix 4 (25% [v/v] MPD, 25% [v/v] PEG 1000, 25% [v/v] PEG 3350). None of the above crystals diffracted in X-ray diffraction experiment.

2.6.9 Structure solution

The X-ray datasets were processed using the dial xia2 pipeline (Winter, 2010) and AIMLESS (Evans and Murshudov, 2013) implemented in CCP4i2 graphical user interface (Potterton et al., 2018). The crystal structures were solved by molecular replacement using PHASER (McCoy et al., 2007). To solve the structure of the ancHMA$_{LVKIE}$/AVR-PikD complex, coordinates of AVR-PikD and a monomer of Pikp-HMA (PDB accession 6fu9) were used. Th obtained structure was refined by iterative rounds of manual rebuilding and refinement with COOT (Emsley et al., 2010) and REFMAC5 (Murshudov et al., 2011), respectively; the structure was accessed with the tools provided in COOT and MolProbity (Chen et al., 2010) and visualised using CCP4MG software (Winn et al., 2011).

2.6.10 Homology modelling

Homology modelling of ancHMA structure in complex with AVR-PikD was built using SWISS-MODEL (Waterhouse et al., 2018) using coordinates of Pikm-HMA–AVR-PikD structure (PDB accession 6g10) as a template.

2.7 Cell death assay

Expression constructs and conditions used for cell death assay are listed in **Table 2.10**. Transient expression in *Nicotiana benthamiana* leaves was conducted as previously described (Bos et al., 2006). Briefly, GV3101::pM90 *Agrobacterium tumefaciens* strains carrying the appropriate expression vectors were mixed and re-suspended in infiltration buffer (10 mM 2-[N-morpholine]-ethanesulfonic acid [MES]; 10 mM $MgCl_2$; and 150 μM acetosyringone, pH 5.6) to a desired density. Upper leaves of 4–5-week-old *Nicotiana benthamiana* plants were used for infiltration. The hypersensitive response (HR), cell death, was scored 5 days after agroinfiltration, unless stated otherwise, using

previously published scale (Segretin et al., 2014) modified to range from 0 (no visible necrosis) to 7 (confluent necrosis).

Table 2.10 List of constructs used in cell death assays

Protein name	Tag	Vector backbone	Concentration (OD600)	Reference
Pikp-1	6×HA (N-term)	pICH77732	0.15	This study
Pikp-1*	6×HA (N-term)	pICH77732	0.15	This study
Pikp-2	3×Myc (C-term)	pCambia	0.15	Maqbool et al., 2015
AVR-PikD	3×FLAG (C-term)	pTRBO	0.15	This study
AVRblb2	3×FLAG (C-term)	pTRBO	0.15	Bozkurt et al., 2011
Pikp-1	6×His/ 3×FLAG (HF) (C-term)	pICH47742	0.4	De la Concepcion et al., 2018
Pikp-1*	6×His/ 3×FLAG (HF) (C-term)	pICH47732	0.4	This study
Pikp-2	6×HA (C-term)	pICH47751	0.4	De la Concepcion et al., 2018
AVR-PikD	4×Myc (N-term)	pICH47732	0.6	De la Concepcion et al., 2018
P19	NA	pCB301	0.1	Win and Kamoun, 2003
Pikm-1	6×His/ 3×FLAG (C-term)	pICH47742	0.4	De la Concepcion et al., 2018
Pikm-1*	6×His/ 3×FLAG (C-term)	pICH47732	0.4	This study
Pikm-2	6×HA (C-term)	pICH47751	0.4	De la Concepcion et al., 2018

*mutants and fusions

NA: not applicable

Chapter 3: The HMA integration into Pik-1 predates the emergence of Oryzinae

3.1 Introduction

Plants have evolved nucleotide-binding leucine rich repeat (NLR) intracellular immune receptors to defend themselves against invading pathogens (Jacob et al., 2013; Kourelis and van der Hoorn, 2018). Although the majority of NLRs bear a broadly conserved modular domain architecture (Duxbury et al., 2020; Jones et al., 2016), a subset contain noncanonical so called integrated domains (IDs) that are thought to mediate recognition of pathogen effectors. Despite the fact that NLR-IDs form anywhere between 3 to 10% of all plant NLRs (Kroj et al., 2016; Sarris et al., 2016), our knowledge about their emergence and subsequent evolution remains sparse.

Several NLR-IDs appear to have lost the ability to autonomously activate defence response (Césari et al., 2014b; Zdrzałek et al., 2020). As a consequence, they often function in pairs, where NLR-ID serves as a sensor detecting pathogen effectors and its partner acts as a helper, or executor, that mediates activation of immune signalling (Adachi et al., 2019b; Bonardi et al., 2011; Feehan et al., 2020). There are now many examples of such NLR pairs, namely RRS1/RPS4 from *Arabidopsis thaliana* (Saucet et al., 2015) and Pik-1/Pik-2 (Ashikawa et al., 2008), Pii-1/Pii-2 (Fujisaki et al., 2017), and RGA4/RGA5 (the *Pia* locus) (Césari et al., 2014b; Okuyama et al., 2011) from rice. These—and many other, not yet characterised, NLR pairs (Bailey et al., 2018)—are encoded by two adjacent genes in a head-to-head orientation. Conceivably, such genetic linkage provides an evolutionary advantage by facilitating co-segregation, coevolution, and transcriptional coregulation of functionally linked genes (Baggs et al., 2017; Griebel et al., 2014). Genetic linkage may also reduce the genetic load caused by autoimmunity, especially for NLR pairs that function though negative regulation (Wu et al., 2017).

Rice Pik-1 and Pik-2 proteins are a well-known pair of CC-NLR receptors. Pik-1 acts as a sensor that binds the AVR-Pik effector of the rice blast fungus via the Pik-1–integrated HMA domain, and Pik-2 is required for activation of immune response upon effector recognition (Maqbool et al., 2015; Zdrzałek et al., 2020). The genetic locus encoding Pik-1 and Pik-2 resistance proteins has been extensively studied

for over a decade. The pair was initially cloned by map-based cloning from Tsuyuake rice (Ashikawa et al., 2008), and has since been shown to occur in allelic variants, which include *Pikp*, *Pikm*, *Piks*, *Pikh*, and *Pik**, to name a few (Costanzo and Jia, 2010; Jia et al., 2009; Wang et al., 2009; Yuan et al., 2011; Zhai et al., 2011). The *Pik* alleles exhibit high levels of polymorphisms, located primarily within its integrated HMA domain (Białas et al., 2018; Costanzo and Jia, 2010; Kanzaki et al., 2012). Numerous studies exploring the genetic diversity of *Pik* revealed that pathogen recognition spectra can be traced to antagonistic molecular interactions of Pik alleles with allelic variants of the cognate AVR-Pik effector (De la Concepcion et al., 2020, 2018; Kanzaki et al., 2012; Li et al., 2019; Longya et al., 2019; Wang et al., 2009). This suggests that these polymorphisms likely emerged as a result of accelerated arms race with the rice blast fungus (Białas et al., 2018; Kanzaki et al., 2012).

As with many other NLR receptor genes (Choi et al., 2016; Christopoulou et al., 2015; Jupe et al., 2012), the *Pik* locus occupies a diverse NLR gene cluster located on the long arm of chromosome 11. This cluster carries signatures of frequent rearrangements evident by presence/absence polymorphisms and gene duplications (Rice Chromosomes 11 and 12 Sequencing Consortia, 2005; Stein et al., 2018; Wang et al., 2009). The *Pik* locus itself also appears to have undergone multiple segmental duplication and rearrangement events that led to emergence of numerous paralogues present across certain rice accessions (Mizuno et al., 2020; Xiahong et al., 2011). Distinct genomic structure of the *Pik* locus in Nipponbare and Kusabue/K60 cultivars represents a well-documented example of haplotype variation that stemmed from genomic rearrangements (Ashikawa et al., 2008; Yuan et al., 2011; Zhai et al., 2011). The N- and K-type *Pik*—named after their parental cultivars, Nipponbare and K60, respectively—constitute intraspecific haplotypes present in a genetic pool of wild and cultivated rice (Zhai et al., 2011). Interestingly, while the K-type receptors mediate resistance against the AVR-Pik effectors, the N-type haplotypes are not known to detect AVR-Pik or any other effectors (Kanzaki et al., 2012; Maqbool et al., 2015; Zhai et al., 2011). This is likely a consequence of indels present in the HMA region of the N-type Pik-1 haplotypes when compared to the HMA domains present in the K-type Pik alleles. Finally, long-term maintenance of haplotype variation at the *Pik* locus might be indicative of balancing selection imposed on this locus, as proposed for a number of NLR genes

(Brabham et al., 2017; Huang et al., 2009; Kanzaki et al., 2012; Karasov et al., 2014; Koenig et al., 2019; Wu et al., 2017).

Despite high levels of polymorphisms, the chromosomal region encompassing the *Pik* locus also displays a certain degree of collinearity across Oryzinae tribe (Stein et al., 2018). In fact, the *Pik* genes have been found in syntenic regions in Oryzinae species, including *Oryza brachyantha* and *Leersia perreri*, suggesting that Pik-1 and Pik-2 comprise an ancient NLR pair that emerged long before rice domestication. Given this rich genetic diversity, the *Pik* genes can serve as a unique experimental system for studying the long-term evolution of NLR receptors, in particular NLR-IDs. In this chapter, I investigated the extent to which the *Pik-1*/*Pik-2* genes are conserved across distantly related grass (Poaceae) species, focusing on two major grass lineages: the PACMAD clade (for Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, and Danthonioideae) comprising of maize, foxtail millet, and sorghum, and the BOP clade (for Bambusoideae, Oryzoideae, and Pooideae) that includes agronomically important crops, such as rice, wheat, and barley. I examined the genomic architecture of the *Pik* locus across selected grass species and explored potential coevolution of Pik-1 and Pik-2 receptors. Furthermore, I used the Pik-1–integrated HMA domain as an evolutionary footprint to unravel the evolutionary history of the Pik receptors and develop hypotheses about the selective forces that shaped their evolution. In particular, I looked at signatures of selection across selected domains of Pik proteins and discovered that the integrated HMA domain stood out by displaying elevated rates of nonsynonymous substitutions relative to synonymous changes. This is indicative of positive selection, which likely reflects a history of evolutionary arms race with pathogen effectors.

3.2 Results and discussion

3.2.1 Pik orthologues are widely present across distantly related grass species

To understand the diversity of the *Pik-1* and *Pik-2* genes across the Poaceae family, I performed a phylogenetic analysis of the entire repertoire of CC-type NLR proteins (predicted NLRome) from eight representative grass species. I used NLR-Parser (Steuernagel et al., 2015) to identify sequences of putative NLRs from publicly available

protein databases of *Zea mays*, *Sorghum bicolor*, *Setaria italica*, *Triticum aestivum*, *Hordeum vulgare*, *Brachypodium distachyon*, *Oryza brachyantha*, and *Oryza sativa*, and, following rigorous filtering steps (described in **Chapter 2**), I compiled a list of 3,062 putative CC-NLRs, amended with known and functionally characterized NLR-type proteins from grasses (**Table 2.2**), including Pikp, Pikm, Piks, Pikh, and Pik* alleles of Pik-1 and Pik-2. Next, I constructed a maximum likelihood phylogenetic tree based on the protein sequences of the NB-ARC domain of the recovered CC-NLRs and discovered that Pik-1 and Pik-2 sequences fell into two phylogenetically unrelated, but well-supported, clades (**Figure A.2.1 A**). Among Pik-1– and Pik-2–related sequences I detected representatives from different, often distantly related, grass species, including members of Pooideae and Panicoideae subfamilies. I propose that the identified clades consist of Pik-1 and Pik-2 orthologues from a diversity of grass species.

In order to determine the topologies within those clades, I performed additional phylogenetic analyses using codon-based sequence alignments of Pik-1 and Pik-2 clade members. Both Pik-1 and Pik-2 phylogenetic trees, calculated using the maximum likelihood method, revealed the relationships within those clades (**Figure A.2.1 B**).

I noted that Pik-2 from *Oryza brachyantha*, the closest homologue of the *Os*Pik-2 alleles, was N-terminally truncated as a result of a 47-bp deletion within its 5'-region (**Figure A.2.2 A**). To determine whether *Oryza brachyantha* population carries a full-length *Pik-2* gene, I genotyped additional *O. brachyantha* accessions (**Figure A.2.2 B**; **Table 2.4**). I successfully amplified and sequenced six additional full-length *ObPik-2* genes, none of which carried the deletion present in the reference genome. I further amplified full-length *ObPik-1* genes from the selected accessions (**Table 2.4**), confirming that the both full-length Pik-2 as well as Pik-1 are present in this species.

Following on these results, I expanded the search of Pik orthologues to additional species, focusing mainly on members of the Oryzoideae subfamily (**Table 2.1**). Using recurrent BLAST searches combined with manual gene annotation and phylogenetic analyses, I was able to identify additional Pik-related NLRs resulting in a total of 41 and 44 Pik-1 and Pik-2 sequences, respectively (**Figure 3.1**). Altogether, the Pik orthologues were widely present across distant grass species. The majority of species within the Oryzinae tribe contained single copies of Pik-1 and Pik-2 per accession, whereas members of the Pooideae and Panicoideae subfamilies frequently encoded multiple Pik-1 or Pik-2

homologues, with wheat carrying as many as nine and ten *Pik-1* and *Pik-2* genes, respectively. Both Pik-1 and Pik-2 from the *Oryza* genus formed two subclades, corresponding to two haplotypes identified at the *Pik* locus: N-type and K-type, named after Nipponbare and Kusabue/K60 rice cultivars, respectively (Ashikawa et al., 2008; Yuan et al., 2011; Zhai et al., 2011) (**Figure A.2.3**). Consistently with previous reports, my phylogenetic analyses demonstrated that the N- and K-type *Pik* genes have been maintained through speciation and co-exist as haplotypes in different *Oryza* species.

To conclude, I discovered that Pik-1 and Pik-2 orthologues are present across a wide range of grasses, including members of the Oryzoideae as well as Pooideae and the Panicoideae subfamilies. I noted that *Pik* genes display marked copy number variations across species. While in the majority of species Pik-1 and Pik-2 are typically encoded by single-copy genes, some, like wheat, occur in expanded phylogenetic clusters. In the following sections, I described various features observed across different Pik orthologues such as genetic linkage of *Pik* genes (**Section 3.2.2**), presence/absence of the Pik-1–integrated HMA domain (**Section 3.2.3**), and the conclusions that can be drawn from these observations.

3.2.2 The *Pik-1*/*Pik-2* gene pair likely emerged early in the evolution of the Poaceae family

3.2.2.1 Genetic linkage of the *Pik* gene pair predates Oryzinae speciation

Pik-1 and Pik-2 constitute a well-known example of functionally and genetically linked NLR pair. In rice the *Pikp-1* and *Pikp-2* genes are located in head-to-head orientation at a single locus on the long arm of chromosome 11, and their coding sequences are separated by only ~2.5-kb-long region (Ashikawa et al., 2008; Yuan et al., 2011). To determine whether this genetic linkage is conserved among Pik orthologues in grasses, I examined the genetic loci of retrieved *Pik-1* and *Pik-2* genes (**Figure 3.2**; **Table A.2.1**). The overwhelming majority of species, in which both genes were present, carried at least one Pik pair with adjacent *Pik-1* and *Pik-2* genes in head-to-head orientation. Although the size of the genes and their intergenic regions varied significantly between species, they exhibited largely conserved gene model. Most of the *Pik-2* orthologues featured one intron in their NB-ARC region while the *Pik-1*

genes typically carried one or—for the genes featuring the HMA domain—two introns. In addition, in species that carry multiple copies of *Pik-1* or *Pik-2*, the copies were typically located in close proximity or, as in wheat, in large NLR-rich gene clusters (**Figure A.2.4**; **Table A.2.2**).

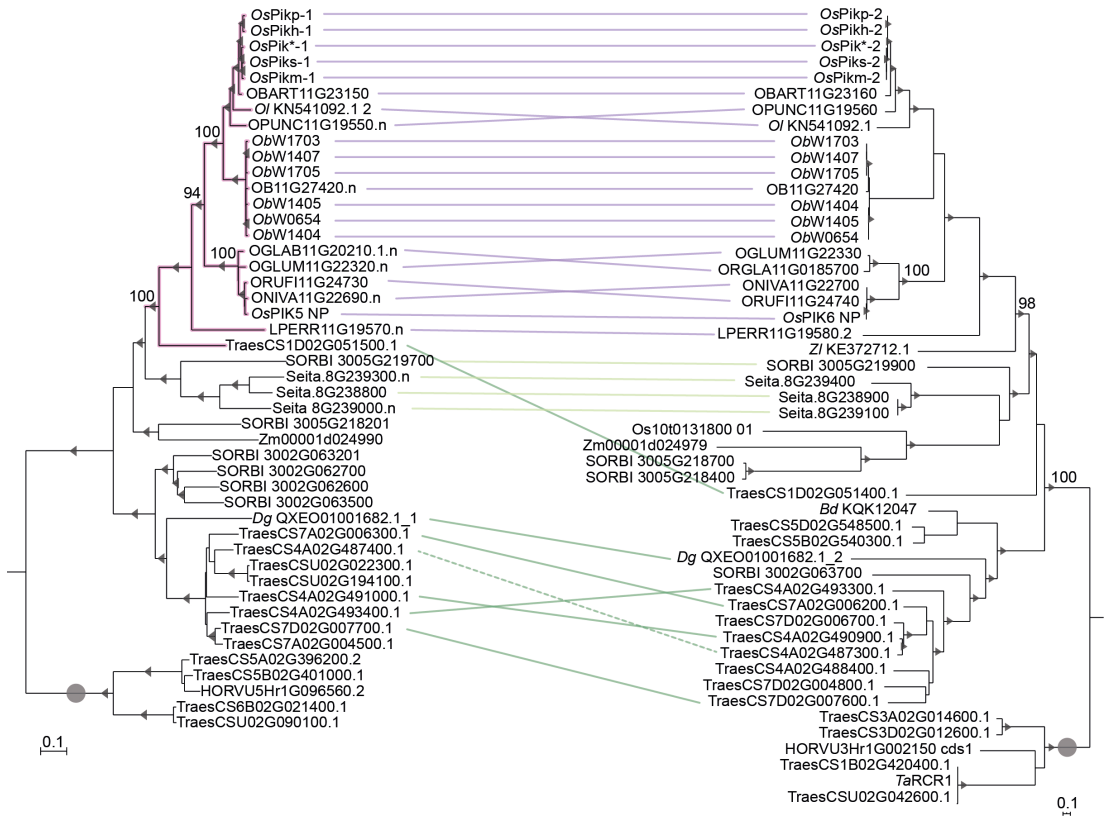


Figure 3.1 The *Pik-1*/*Pik-2* orthologues are distributed across diverse species of grasses

The maximum likelihood (ML) phylogenetic trees of Pik-1 (left) and Pik-2 (right) homologues. The trees were calculated from 927- and 1239-nucleotides-long codon-based alignments, respectively, using RAxML v8.2.11 (Stamatakis, 2014), 1000 bootstrap method (Felsenstein, 1985), and GTRGAMMA substitution model (Tavaré, 1986). Best ML trees were manually rooted using the selected clades (marked with grey circle) as outgroups. The bootstrap values above 70 are indicated with grey triangles at the base of respective clades; the support values for the relevant nodes are depicted with numbers. The scale bars indicate the evolutionary distance based on nucleotide substitution rate. The HMA integration clade is shown in pink. Genetically linked genes are linked with lines, with colours indicating plant subfamily: Oryzoideae (purple), Poodieae (dark green), and Panicoideae (light green); the continuous lines represent linkage in head-to-head orientation, the dashed line indicates linkage in tail-to-tail orientation. The interactive trees are publicly available at: https://itol.embl.de/tree/14915519290329341598279392 for Pik-1 and https://itol.embl.de/tree/14915519290161451596745134 for Pik-2.

Pik-1 and *Pik-2* occupy a phylogenetically heterogeneous NLR gene cluster that shows notable collinearity across genomes of wild rice species (Mizuno et al., 2020; Stein et al., 2018). Here, I identified genetically linked Pik orthologues in more distantly related species, including wheat (*Triticum aestivum*) and foxtail millet (*Setaria italica*), demonstrating that the conservation of the *Pik* locus goes beyond Oryzoideae subfamily. These results suggest that Pik-1 and Pik-2 are encoded by a pair of ancient genes that emerged early in Poaceae evolution, presumably before the divergence of the BOP and PACMAD clades, estimated for anywhere between 50 to 100 MYA (Hodkinson, 2018). Since then the genes have been maintained through speciation of the Poaceae family as a genetically linked pair and can now be found in a wide range of grass species.

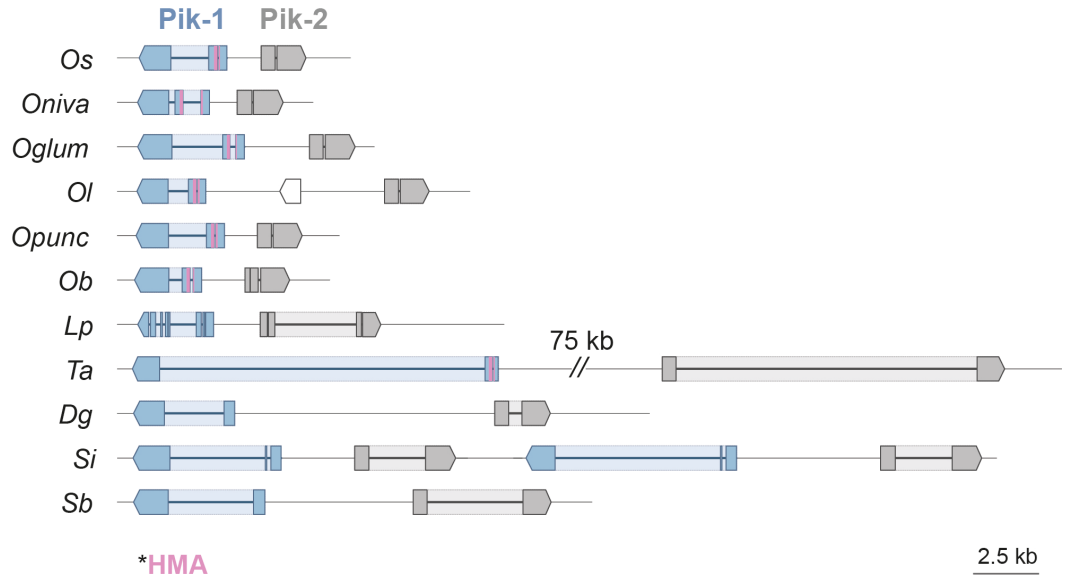


Figure 3.2 Many of the *Pik-1* and *Pik-2* genes in grasses are genetically linked in head-to-head orientation

The schematic of the *Pik* locus in selected species. The schematic gene models of *Pik-1* (blue) and *Pik-2* (grey) are shown. The integrated HMA domain is marked with pink. The coordinates of the regions presented in this figure are summarised in **Table A.2.1**. *Os*: *O. sativa*, *Oniva*: *O. nivara*, *Oglum*: *O. glumaepatula*, *Ol*: *O. longistaminata*, *Opunc*: *O. punctata*, *Ob*: *O. brachyantha*, *Lp*: *L. perrieri*, *Ta*: *T. aestivum*, *Dg*: *D. glomerata*, *Si*: *S. italica*, *Sb*: *S. bicolor*.

3.2.2.2 Genetically linked Pik-1 and Pik-2 have a shared evolutionary history

Given that genomic rearrangements have been reported at the *Pik* locus (Mizuno et al., 2020; Stein et al., 2018), one cannot exclude the possibility that genetic linkage of the *Pik-1*/*Pik-2* pair emerged more than once and is a remnant of rearrangement events. I reasoned that if the gene pair have remained genetically linked over a long evolutionary period, then they should have the same molecular age (as determined by the rate of synonymous substitutions or d_S). To gain insights into the evolutionary dynamics between genetically linked Pik-1 and Pik-2 receptors, I compared d_S rates of linked *Pik* genes. For the analysis, I selected representative Pik-1 and Pik-2 NLRs that are genetically linked in head-to-head orientation from each species. *Lp*Pik orthologues were excluded from the analysis due to their unusual gene models (**Figure 3.2**). Next, I assessed rates of synonymous substitutions (d_S) within the coding sequences of the NB-ARC domain of the selected genes using the method of Yang and Nielsen (2000). This method calculates substitution rates based on pairwise comparison between two given sequences. The rates were calculated separately for Pik-1 and Pik-2 and cross-referenced such that the pairwise values for Pik-1 were compared to the respective values for cognate Pik-2 (**Table A.2.3**). The comparisons revealed strong positive correlation of d_S rates ($R^2 = 0.87$) (**Figure 3.3 A, C**) that was significantly higher than observed by chance, as calculated from random Pik-1–Pik-2 cross-referencing (**Figure A.2.5 A**). The comparisons between sequences from different subfamilies overall showed the weakest correlation, with the corresponding d_S values often displaying less than 70% similarity, whereas genes from the same taxonomic rank typically showed higher d_S similarity. Simultaneously, I calculated and compared the rates of nonsynonymous substitutions (d_N) among the NB-ARC domains of Pik-1 and Pik-2 sequences, and also discovered significant correlation of d_N rates ($R^2 = 0.88$) across Pik pairs (**Figure 3.3 B, C**; **Figure A.2.5 B**).

To conclude, the correlation of rates of synonymous substitutions (d_S) provides independent evidence that *Pik-1* and *Pik-2* have been coevolving, sharing a similar evolutionary history for the last 50 to 100 million years—before the split of the BOP and PACMAD clades (Hodkinson, 2018).

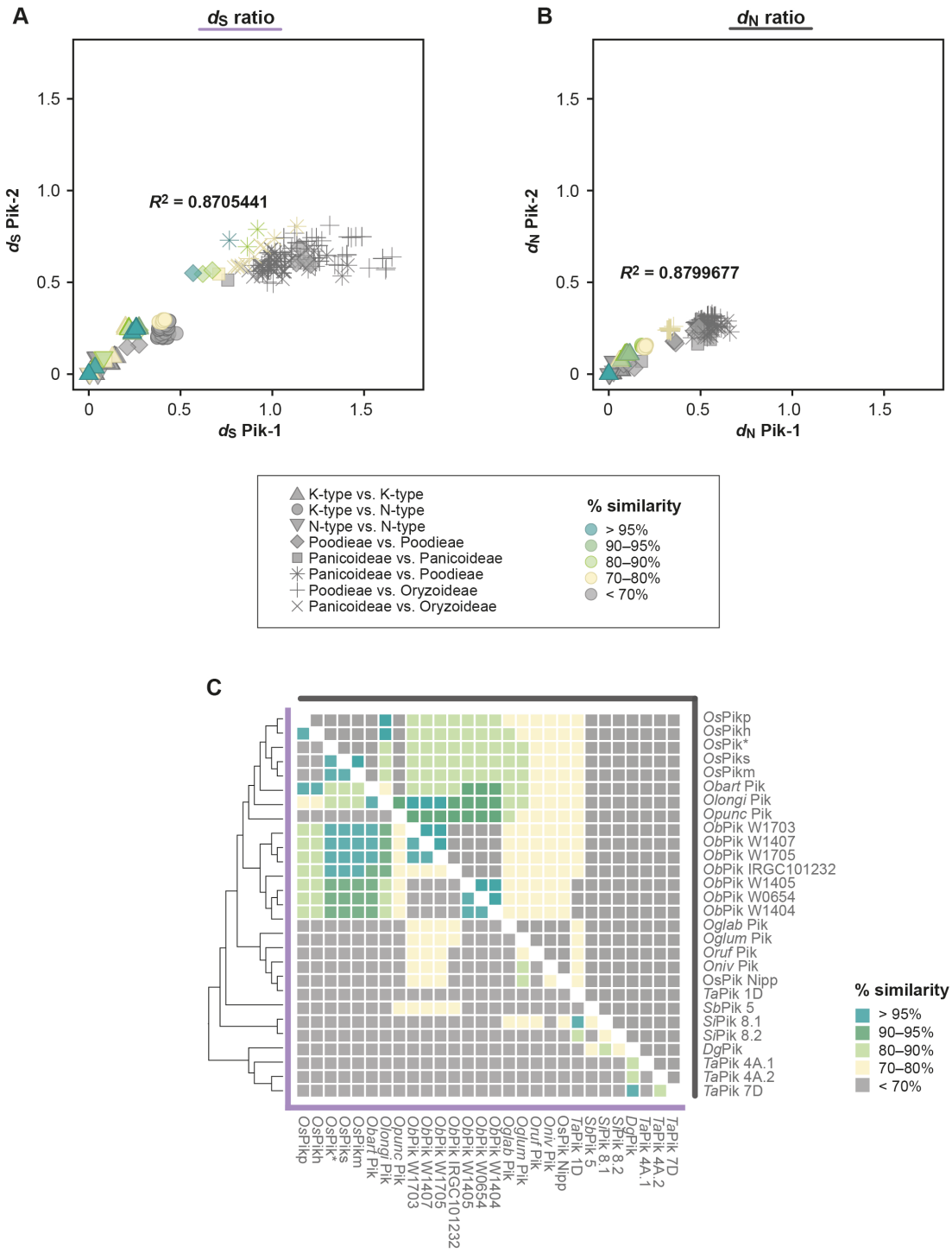


Figure 3.3 Genetically linked Pik-1 and Pik-2 show signatures of shared evolutionary history

Comparisons of pairwise d_S (**A**) and d_N (**B**) rates calculated for the Pik-1 and Pik-2 receptors. The synonymous (d_S) and nonsynonymous (d_N) rates were calculated using Yang and Nielsen method (Yang and Nielsen, 2000) based on 972- and 1269-nt-long codon-based alignments of the »

Figure 3.3 Genetically linked Pik-1 and Pik-2 show signatures of shared evolutionary history (continued)

NB-ARC domains of Pik-1 and Pik-2, respectively; only positions that showed over 70% coverage across the alignment were used for the analysis. The comparisons were categorised to reflect species divergence (shapes) and colour-coded to illustrate percentage similarity of the values of d_S or d_N (% similarity). Coefficient of determination (R^2) was calculated for each dataset using R v3.6.3 package. (**C**) The pairwise comparisons of d_S (lower-left panel, marked with purple line) and d_N (upper-right panel, marked with grey line) rates presented as a heatmap. The comparisons were ordered based on the Pik-1 phylogenetic relationship, shown on the left. The list of genes used for the pairwise comparisons is summarised in **Table A.2.3**.

3.2.3 The HMA integration predates the emergence of Oryzinae

To better understand the evolutionary history of the Pik-1 sensor NLR, I looked for signatures of the HMA integration among its orthologues. Remarkably, I found that the HMA domain was not always present in the *Pik-1* genes. The HMA-containing homologues clustered into a single well-supported clade (herein called the integration clade), whereas the sequences without the integration were located at the branches basal to the integration clade (**Figure 3.1**). With the exception of *Lp*Pik-1 (*Leersia perrieri*), all members of the integration clade carried the HMA domain in the same position—between the CC and NB-ARC domains of the Pik-1 receptor—and featured an intron within the HMA (**Figure 3.2**). This indicates that these integrated HMAs were most likely derived from a single integration event.

Leveraging this information, I generated a multiple sequence alignment of selected Pik-1 orthologues to estimate the HMA integration position (**Figure A.2.6**). In particular, I focused on comparison of representative members of the integration clade and their closest relatives from *Setaria italica* and *Sorghum bicolar* lacking the HMA domain. Based on this comparison, I concluded that the integration site most likely falls between the KLL residues at the N-terminus and KTV at the C-terminus (corresponding to residues 161–163 and 284–286 of Pikp-1); however, the exact boundaries of the integration might be slightly different, given the high sequence divergence near the integration site among more distantly related orthologues. Interestingly, the integration site encompasses a wider region than that of functionally characterised HMA domains (De la Concepcion et al., 2020, 2018; Maqbool et al., 2015), with around 20 additional amino acids (23 and 21 in Pikp-1) on each side of the annotated HMA domain. It's possible that those residues

act as flexible linkers that help to accommodate the HMA domain within the NLR three-dimensional structure and ensure proper domain cooperation.

To estimate when Pik-1 acquired the HMA domain, I compared the phylogeny of plant species included in this study with the presence/absence of the HMA as determined by the Pik-1 phylogenetic tree (**Figure 3.4**). I found that orthologues from the *Oryza* genus accounted for the majority of Pik-1–HMA fusions. In fact, all orthologues of Pik-1 in *Oryza* have the HMA domain. This suggests that the integration unambiguously predates the speciation of this genus. Although I failed to detect a full-length HMA integration in *L. perrieri*, *Lp*Pik-1 carries at least 15 amino acids characteristic of the HMA integration site (**Figure A.2.6**), indicating that the fusion probably occurred before the speciation of Oryzinae and was subsequently lost in *Leersia perrieri*. In contrast, the vast majority of examined Pik-1 from the Poodieae and Panicoideae subfamilies lack the HMA domain. The only integration in these taxonomic groups was detected in one, out of total of nine, Pik-1 paralogue of wheat. This observation may indicate that the Pik-1–HMA fusion emerged prior to radiation of the BOP and PACMAD clades, 100–50 MYA (Hodkinson, 2018). In this scenario, Pik-1 acquired the HMA following gene duplication, then while both of those ancient paralogues have been maintained to present day in wheat, only one of them has been retained in other species, leading to HMA presence/absence polymorphisms observed in grasses. However, it is also possible that the integration occurred much later and that the newly emerged Pik-1–HMA gene transferred to wheat through introgression from rice progenitors. Taken together, these results clearly indicate that HMA integration predates diversification of Oryzinae, with the possibility of it being much more ancient than Oryzinae itself. Future research should address this question, for example by comparing the molecular age of the *Pik-1* genes to the divergence dates of the species in question.

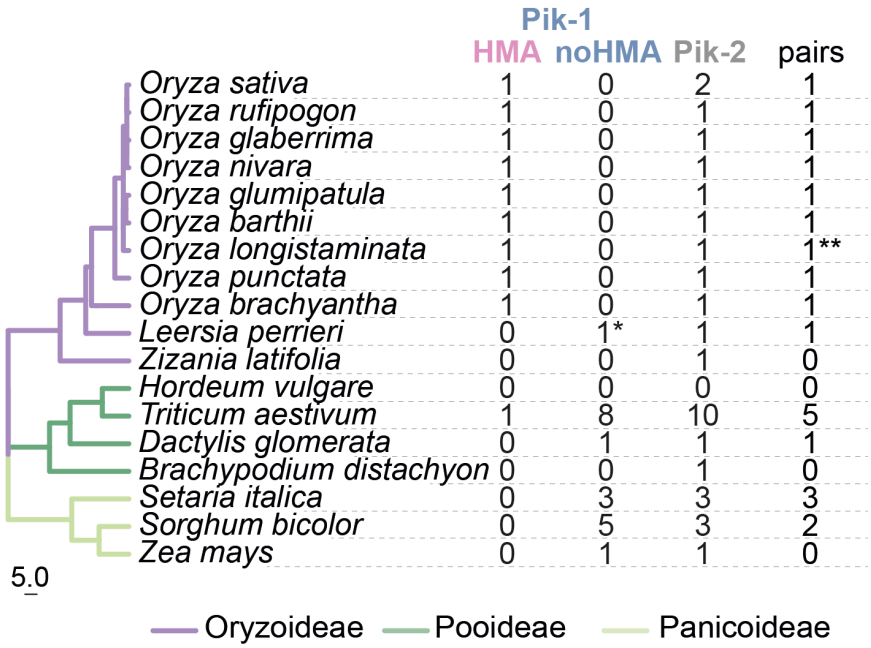


Figure 3.4 The HMA integration into Pik-1 predates the emergence of Oryzinae

Summary of the Pik-1 and Pik-2 homologues identified from different plant species included in this study. The phylogenetic tree was generated using the TimeTree tool (Kumar et al., 2017). The number of pairs correspond to the number of *Pik-1*/*Pik-2* genes in head-to-head orientation separated by intergenic region of various length. (**) The species harbours a truncated gene between *Pik-1* and *Pik-2*. (*) The species has likely lost the HMA.

3.2.4 The integrated HMA domain carries signatures of positive selection

3.2.4.1 The Pik-1–integrated HMA domain displays elevated d_N/d_S ratio indicative of positive selection

In rice, the Pik-1–integrated HMA domain exhibits higher levels of polymorphisms compared with canonical domains of Pik-1 and Pik-2 (Costanzo and Jia, 2010; Kanzaki et al., 2012), which implies that the HMA is under positive selection. To characterise the selection pressures underlying HMA diversification following the integration, I examined molecular signatures of selection within the integration clade. Wheat Pik-1–HMA homologue was excluded from the analysis due to extreme sequence divergence relative to *Oryza* orthologues, which precluded generating reliable sequence alignment. Based on phylogenetic relationship, the remaining sequences were assigned into K- and N-type sequences and analysed separately. To test for molecular signatures of selection, I calculated rates of synonymous (d_S) and nonsynonymous (d_N) substitutions across the

coding sequences of the HMA domain. I discovered that d_N was greater than d_S in 96 out of 115 pairwise sequence comparisons (86/105 for K- and 10/10 for N-type HMAs) (**Figure 3.5 A, C–D**), providing evidence that strong positive selection has indeed acted on the integrated HMA domain. A total of 19 of those cases displayed $d_S = 0$, all of which involved closely related sequences. To test if this phenomenon was specific to the HMA domain, I further calculated d_N and d_S rates for the NB-ARC domain of the same set of genes and discovered that out of total of 115 pairwise sequence comparisons only nine displayed d_N greater than d_S (**Figure 3.5 B, C–D**); however, all of these showed $d_S = 0$, and were therefore inconclusive. A comparison of the d_N and d_S rates between the HMA and NB-ARC domains, further highlighted the elevated rates of nonsynonymous substitutions within the integrated HMA domain relative to the NB-ARC (**Figure 3.5 E–F**).

Overall, these results demonstrate that the integrated HMA exhibits strong signatures of positive selection, in sharp contrast to the NB-ARC domain. I hypothesize that this strong HMA diversification is driven by antagonistic molecular interaction with pathogen effectors, such as AVR-Pik–related effectors from *Magnaporthe oryzae*, which have been shown to interact with a number of K-type Pik-1–integrated HMAs from rice (De la Concepcion et al., 2020, 2018; Maqbool et al., 2015). Notably, the Nipponbare rice cultivar, which encodes N-type Pik, is susceptible to *Magnaporthe oryzae* strains carrying the AVR-Pik effector alleles (Kanzaki et al., 2012; Maqbool et al., 2015; Zhai et al., 2011), suggesting that the mechanisms underpinning the diversification are different in the N- and K-type receptors.

To date, no resistance has been mapped to the *Pik* locus of the Nipponbare cultivar of rice, inviting the question about the determinants of sequence diversifications of the N-type Pik-1–integrated HMA domains. It's interesting to speculate that, given marked sequence polymorphism compared with K-type HMAs, N-type *Pik* lost the ability to recognise pathogen effectors and is therefore under relaxed selection. However, in this scenario one would expect signatures of relaxed selection distributed somewhat evenly across the whole length of the gene, which does not seem to be the case—unlike the HMA, the NB-ARC domain of the N-type genes does not display elevated rates of d_N/d_S (**Figure 3.5 B, C**). Whether N-type *Pik* confer resistance to different, yet unknown,

effectors, and if HMA diversification in those genes led to sub- or neofunctionalization remains to be explored.

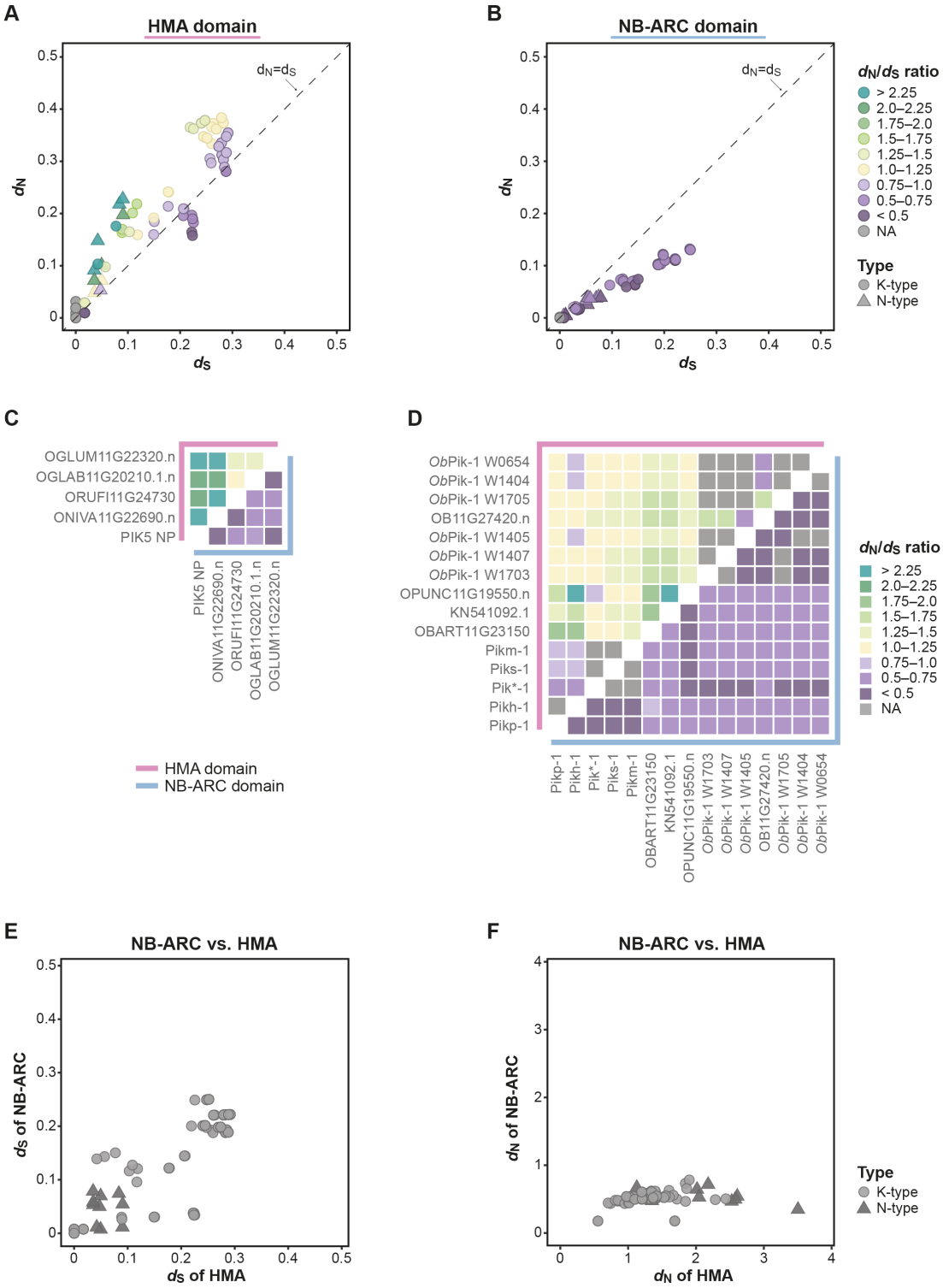


Figure 3.5 The integrated HMA domain experiences elevated rates of d_N/d_S compared with the NB-ARC domain of Pik-1 (continued on the next page)

Figure 3.5 The integrated HMA domain experiences elevated rates of d_N/d_S compared with the NB-ARC domain of Pik-1 (continued)

(**A–B**) Pairwise comparison of nucleotide substitution rates within the HMA integration clade for the HMA (**A**) and NB-ARC (**B**) domains were calculated using Yang and Nielsen method (Yang and Nielsen, 2000). The diagonal line (dashed) indicates $d_N = d_S$ meaning neutral selection; points above this line indicate positive (diversifying) selection with $d_N/d_S > 1.0$, whereas points below represent negative (purifying) selection with $d_N/d_S < 1.0$. The points are colour-coded to indicate d_N/d_S ratio; NA: the ratio was not calculated because $d_S = 0$. The pairwise comparisons were performed for the K-type (circles) and N-type (triangles) Pik-1 sequences separately. (**C–D**) To highlight the differences between the d_N/d_S ratios for the HMA and NB-ARC domains the rates were plotted as heatmaps corresponding to N- (**C**) and K-type (**D**) Pik-1 sequences. (**E–F**) Pairwise comparison of d_N (**E**) and d_S (**F**) rates between the HMA and NB-ARC domains of Pik-1.

3.2.4.2 Identification of amino acids within the integrated HMA domain that are likely under positive selection

Positive selection typically acts only on particular amino acids within the protein. Therefore, I aimed to detect individual sites that experienced positive selection within the integrated HMA domain using different codon substitution models for heterogeneous selection at amino acid sites. For the analysis, I focused on the K-type HMA domain, which likely coevolved with the AVR-Pik effectors. To capture more genetic diversity of the K-type Pik-1–integrated HMAs, I first genotyped additional wild rice species for presence of the integration. Based on the available *Pik-1* sequences from *O. sativa* and *O. brachyantha* I designed primers flanking the HMA interaction, and used PCR amplification to detect the HMA domain in wild rice accessions. I was able to detect the HMA integration in 21 accessions from 13 species (**Table A.2.4**); ten of those showed sufficient coverage across the entire functional region of the HMA and were included in the analysis (**Figure 3.6 A**). To detect patterns of selection within the integrated HMA, I applied three pairs of maximum likelihood (ML) models of codon substitution: M3/M0, M2/M1, and M8/M7 (Yang et al., 2000). All three pairwise tests confirmed that the selection varied across different sites within the HMA (**Figure 3.6 B**). As indicated by posterior probabilities, the tests identified signatures of positive selection at the same sites, although with different confidence levels (**Figure 3.6 C**). Many of those were located at well-characterised effector-binding interfaces, including positions 228 and 261, which have been shown to mediate interactions with AVR-Pik in Pikp- and Pikm-1, respectively

(De la Concepcion et al., 2018; Maqbool et al., 2015). On average, residues located in the unstructured regions of the HMA (loops) showed the highest probabilities of positive selection. To date, however, the function of these residues remains elusive.

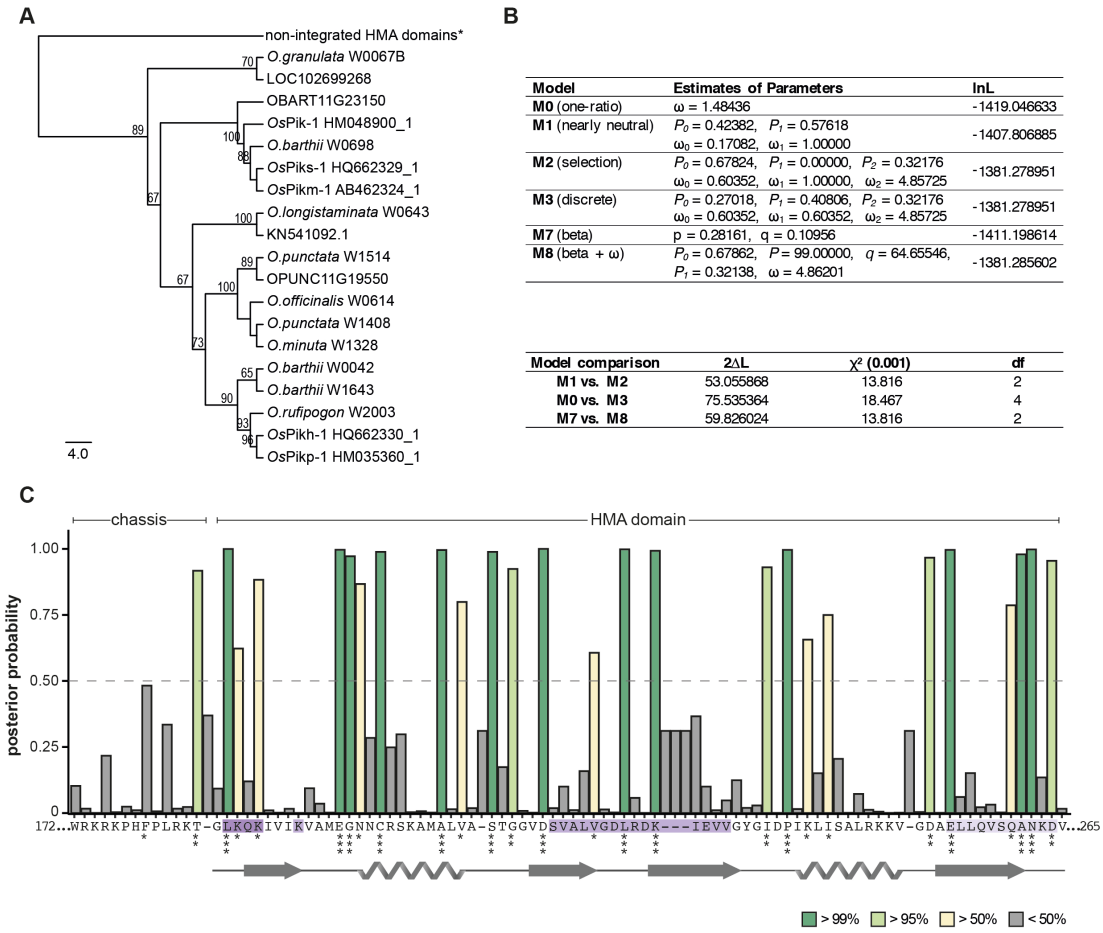


Figure 3.6 Residues within the integrated HMA domain are likely to have experienced positive selection

(**A**) The neighbour joining tree of the HMA domain calculated using the JTT substitution model (Jones et al., 1992) and bootstrap method with 100 iterations test (Felsenstein, 1985). Alignment of 98 amino acids of integrated HMAs was generated with MUSCLE (Edgar, 2004). Bootstrap values above 65% were shown at the base of respective clades. The scale bar marks the evolutionary distance based on number of base substitutions per site. (*) A branch corresponding to non-integrated HMAs was manually added to the tree to indicate an outgroup, which was used for tree calculation but not for calculating the selection probabilities. The entire tree is presented in **Figure 4.1**. (**B**) Results from codon substitution models for heterogeneous selection at amino acid sites (upper panel) and likelihood ratio test (bottom panel). (**C**) Posterior probabilities for site classes estimated under the beta & ω (M8) model inferred using Bayes empirical Bayes (BEB) (Yang, 2005). The amino acids with higher values of posterior probability are more likely to be under positive selection. The stars indicate potentially positively selected sites: (*) PP>50%, (**) PP>95%, (***) PP>99%. The amino acid sequence and the protein model showed below the plot correspond to Pikp-1. The effector-interaction interfaces are marked in shades of purple.

As a control for these analyses, I performed the same tests on the NB-ARC domain of the K-type Pik-1 sequences. The discrete model suggested that d_N/d_S varied among different amino acid sites (**Figure A.2.7**); nevertheless, other tests failed to detect signatures of positive selection. Based on these results, I concluded that, consistent with my previous findings, there is no evidence that the NB-ARC domain of Pik-1 as a whole is subject to positive selection, unlike the HMA domain.

3.3 Conclusions

In this chapter, I uncovered the genetic diversity of the Pik-1 and Pik-2 NLRs in grasses. I discovered that *Pik* genes are present in a wide panel of Poeaceae species, including members of the Oryzoideae as well as the Pooideae and the Panicoideae subfamilies. In the majority of cases the *Pik* genes are genetically linked in head-to-head orientation and their homologues harbour similar gene model. In addition, the Pik-1 and Pik-2 homologues show hallmarks of shared evolutionary history, which suggests that they derived from a single ancestral pair that emerged before the radiation of the BOP and PACMAD clades. By looking at the presence/absence of the HMA domain within the Pik-1 orthologues, I was able to determine that the HMA integration likely predates speciation of Oryzinae, with the possibility of it being much more ancient than Oryzinae itself. Regardless of the exact integration date, it appears that the *Pik-1*/*Pik-2* pair existed prior the integration, which contradicts current dogma that the recruitment of a helper NLR is most likely a consequence of an emergence of an NLR sensor with extraneous domain. These findings shed new light on our understanding of evolution and emergence of NLR-IDs. Finally, I confirmed that the HMA domain of Pik-1 stands out by being under strong diversifying selection, likely driven by interactions with pathogen effectors following the integration. Future studies should investigate whether Pik orthologues, with or without the HMA integration, functionally and mechanistically overlap to mediate resistance against the rice blast fungus or other pathogens.

Chapter 4: Pikp-1 evolved towards high affinity binding to the AVR-PikD effector from the rice blast fungus

4.1 Introduction

Nucleotide-binding domain leucine-rich repeat-containing (NLR) proteins constitute an ancient class of intracellular immune receptors that confer innate immunity in plants and animals (Dodds and Rathjen, 2010; Jones et al., 2016). The majority of plant NLRs share a conserved modular architecture, consisting of varying N-terminal, NB-ARC, and LRR domains (Takken and Goverse, 2012). A subset of them, however, carry unconventional domains known as integrated domains (IDs) (Kroj et al., 2016; Sarris et al., 2016). Many IDs are thought to have derived from effector-associated host proteins—also called effector operative targets—and act as baits for effector recognition within NLRs.

To date, the molecular basis of effector recognition and activation of immune response has been studied for only a handful of NLR-IDs. Studies of Arabidopsis RRS1-R revealed that its integrated WRKY domain can recognise AvrRps4 from *Pseudomonas syringae* through binding and PopP2 from *Ralstonia solanacearum* by acting as a substrate for its enzymatic activity (Heidrich et al., 2013; Sarris et al., 2015). The HMA domains of the RGA5 and Pik-1 receptors of rice mediate recognition of the blast effectors AVR-Pia/AVR1-CO39 and AVR-Pik, respectively, via direct binding (César et al., 2013; De la Concepcion et al., 2020, 2018; Guo et al., 2018; Maqbool et al., 2015; Ortiz et al., 2017). In the case of RGA5, simultaneous interaction with the RGA5-HMA domain as well as other parts of the RGA5 receptor is seemingly required for immunity activation (Ortiz et al., 2017). Similarly, effector binding is not sufficient to activate Pik-mediated immune response. Studies dissecting recognition specificity of Pik-1 alleles demonstrated that the HMA domain of the Pikp-1 allele (Pikp-HMA) is able to bind three allelic variants of the AVR-Pik effector—AVR-PikA, AVR-PikD, and AVR-PikE—yet, somewhat surprisingly, only interaction with AVR-PikD leads to immune response and resistance (De la Concepcion et al., 2018; Kanzaki et al., 2012; Maqbool et al., 2015). This discrepancy has been linked to the binding affinity between the HMA and the effector;

interaction with the AVR-PikA and AVR-PikE alleles is at least ten times weaker than that of AVR-PikD, indicating that strong binding to the effector is essential for Pik-mediated immune response (De la Concepcion et al., 2018; Maqbool et al., 2015). The Pikm-1 and Pikh-1 alleles, on the contrary, exhibit much tighter interaction with these effectors, thus recognising and conferring resistance to all three of them (De la Concepcion et al., 2020, 2018). These discoveries have been successfully utilised for structure-guided receptor engineering that resulted in the Pikp-1 mutant with broader effector recognition spectrum (De la Concepcion et al., 2019).

Undoubtedly, new advancements in biophysics and biochemistry have accelerated research on NLR-IDs providing invaluable information that could be used to dissect the mechanisms of their function and help in generating receptors with new recognition specifies, as illustrated by research on the Pik-1 receptors (De la Concepcion et al., 2019). The mechanisms underlying effector recognition by the Pik-1–integrated HMA domain have been well-characterised using plethora of methods, including structural biology, in vitro interaction studies, in planta co-immunoprecipitation, cell death assays in *Nicotiana benthamiana* leaves, rice infection assays, and others (De la Concepcion et al., 2020, 2019, 2018; Kanzaki et al., 2012; Maqbool et al., 2015). However, despite these advancements, there are still gaps in our understating of the system, in particular in the context of Pik evolution. For instance, what has the evolutionary trajectory of the HMA domain been following its integration into an NLR protein? Can we identify the adaptive mutations that enabled recognition of the AVR-Pik effectors and estimate their emergence time? Are there any other selection pressures that have shaped HMA evolution besides effector binding?

In recent years, leveraging evolutionary information to dissect protein function has gained growing interest (Dean and Thornton, 2007; Harms and Thornton, 2013; Thornton, 2004). After decades of parallel research, molecular evolution and biochemistry are starting to be used in conjunction to unravel the molecular basis of protein function within an evolutionary framework. An understanding of how proteins came to be the way they are—how evolution has shaped protein properties and how those properties have affected protein evolutionary trajectories—can inform mechanistic understanding of protein function. After all, protein evolution has been a massive billion-years-old experiment and it seems only natural to seek to understand this experiment.

One approach to study molecular evolution is ancestral protein reconstruction, which in combination with biochemistry has been successfully applied for a great number of proteins, including enzymes (Kaltenbach et al., 2018; Nguyen et al., 2017), haemoglobin (Pillai et al., 2020), antigen receptors (Rouet et al., 2017; Voordeckers et al., 2012), hormone receptors (Bridgham, 2006; Ortlund et al., 2007), influenza virus coat protein (Su et al., 2015), and more recently plant pathogen effectors (Dong et al., 2014; Tanaka et al., 2019) (**Figure 1.1**).

Despite the wealth of knowledge about mechanisms governing effector recognition by the Pik-1–integrated HMA domain, we know very little about its evolutionary history. The research summarised herein stems from the model in which Pik-1 and AVR-Pik have evolved through an arms race—a tight interplay shaping the evolution of both parties—as they both carry signatures of rapid evolution (**Figure 1.2**) (Białas et al., 2018; Kanzaki et al., 2012). Allelic variants of AVR-Pik carry only five amino acid replacements, but no synonymous changes (Huang et al., 2014; Kanzaki et al., 2012; Li et al., 2019; Longya et al., 2019; Yoshida et al., 2009). Remarkably, all nonsynonymous mutations map to regions located at the binding interface with the HMA domain of Pik-1, indicating the adaptive nature of those polymorphisms (De la Concepcion et al., 2020, 2018; Longya et al., 2019; Maqbool et al., 2015). Conversely, the integrated HMA is the most polymorphic region among rice Pik-1 receptors (Białas et al., 2018; Costanzo and Jia, 2010; Zhai et al., 2014) and shows strong signatures of positive selection as demonstrated in the previous chapter (**Chapter 3**).

In this chapter, I aimed to test hypotheses about adaptive evolution of the Pikp-1–integrated HMA domain (Pikp-HMA) and bridge the gap between mechanistic and evolutionary understanding of the Pik immune system. The central question I pursue in this chapter is whether historical emergence of the HMA integration and its diversification within Pikp-1 are a direct adaptation to the most ancient of the AVR-Pik alleles, AVR-PikD, and to what extent the evolutionary trajectory of Pikp-HMA has been driven by this effector. I hypothesised that the evolutionary path of Pikp-HMA must have been shaped by effector recognition but also by constraints posed by other components of the Pikp complex, such as other domains of the Pikp-1 or Pikp-2 proteins. To challenge these questions, I performed ancestral sequence reconstruction coupled with resurrection and functional characterisation of the extinct ancestral HMA. Using established

biophysics and biochemistry methods, such as crystallography and co-immunoprecipitation, I elucidated how historical mutations have altered the HMA properties in the context of effector recognition and immune response activation, and identified key molecular adaptations towards AVR-PikD binding. This work provides an example of a robust evolutionary framework for studying mechanisms of protein adaptation.

4.2 Results and Discussion

4.2.1 Ancestral sequence recontraction of the Pikp-1–integrated HMA domain

To understand the evolutionary trajectory of the Pik-1–integrated HMA domain, I reconstructed ancestral HMA sequences at the early stages of rice speciation based on representative phylogenetic tree of the K-type integrated HMA domains. As an outgroup I selected sequences of HMA domains of the heavy metal–associated isoprenylated plant proteins (HIPPs)—hereafter called non-integrated HMAs—from *Oryza sativa* and *Oryza brachyantha* (De Abreu-Neto et al., 2013). Determining fine phylogenetic relationships using short sequences of rapidly evolving genes can be particularly challenging as characterised by low bootstrap support for some of the branches. Thus, to perform the reconstruction, I first tested different phylogenetic methods and focused on nodes that were well-supported in both the neighbor noining and maximum likelihood phylogenies generated from a codon-base alignment (**Figure A.3.1**). Selected nodes were located at the base of the phylogenies and corresponded to the early stages of rice diversification. Next, I performed the ancestral sequence prediction based on protein sequence alignment, using FastML software (Ashkenazy et al., 2012), which has been previously shown to infer ancestral sequences with high accuracy (Randall et al., 2016). Multiple reconstructions yielded multiple plausible ancestral HMA (ancHMA) variants. To ensure maximum accuracy and to reduce the possibility of incorrect prediction, I selected six representative sequences for further studies (**Figure 4.1**). The selected ancestral sequences were inferred with strong support; the mean and median posterior probabilities (PP) for all nodes were greater than 93% and 99%, respectively; over 78%,

83%, 79%, 80%, and 90% of sites within I-N2, I-N6, II-N11, II-N12, III-N11, and III-N12 variants, respectively, had PP greater that 95%.

Comparative sequence analysis of the ancestral HMAs to the modern HMA domains revealed polymorphism hotspots (**Figure A.3.2**), many of which were located at the effector-binding interfaces (De la Concepcion et al., 2018). This observation suggests that the ability to perceive pathogen effectors has likely been under positive selection. Moreover, there is marked sequence diversity located in the heavy metal-binding motif MxCxxC, characteristic for members of heavy metal–binding protein family (De Abreu-Neto et al., 2013; DeSilva et al., 2002). The motif is degenerated in the panel of present-day Pik HMAs, such as Pikp-HMA, whereas inferred ancHMA variants contained potentially active MxSxxC or MxSxxS motifs (**Figure 4.1**) (Banci et al., 2008). Although it remains to be determined whether the ancHMAs—or any of the non-integrated HMA-containing proteins of rice for that matter—can interact with heavy metals, it is tempting to speculate that following the integration the HMA domain lost the ability to bind metals, either as a result of neutral drift or because it interfered with either effector recognition or immune response activation. Both the phylogenetic analysis and the sequence comparison highlight that following the integration the HMA domains of Pik-1 underwent different evolutionary pathways from their non-integrated relatives, presumably due to different evolutionary constraints.

4.2.2 The HMA domain of Pikp-1 evolved from weak to strong interaction with the AVR-PikD effector

Building on the fact that strong binding to the effector is required for Pik-mediated immune response (De la Concepcion et al., 2020, 2019, 2018; Maqbool et al., 2015), I hypothesised that the HMA domain of Pikp-1 evolved towards high binding affinity to the AVR-PikD effector from a weaker ancestral state. To test this hypothesis, I resurrected the six ancHMA variants by synthesising their predicted sequences and incorporating them into the Pikp-1 receptor—generating Pikp-1:I-N2, Pikp-1:I-N6, Pikp-1:II-N11, Pikp-1:II-N12, Pikp-1:III-N11, and Pikp-1:III-N12 fusions—and tested their association with AVR-PikD in in planta co-immunoprecipitation (co-IP) experiments (**Figure A.3.3 A**). The western blot analysis revealed that the ancHMA variants exhibited a range of association strength with AVR-PikD (**Figure A.3.3 B**). Even so, in every case

the association was consistently weaker than that of the present-day Pikp-1, indicating that binding strength has indeed changed over the course of the Pikp-HMA evolutionary history. For further studies, I selected the I-N2 ancHMA variant, hereafter called ancHMA, which is the last common ancestor of Pik*-1, Pikp-1, Pikh-1, Piks-1, and Pikm-1 alleles of rice.

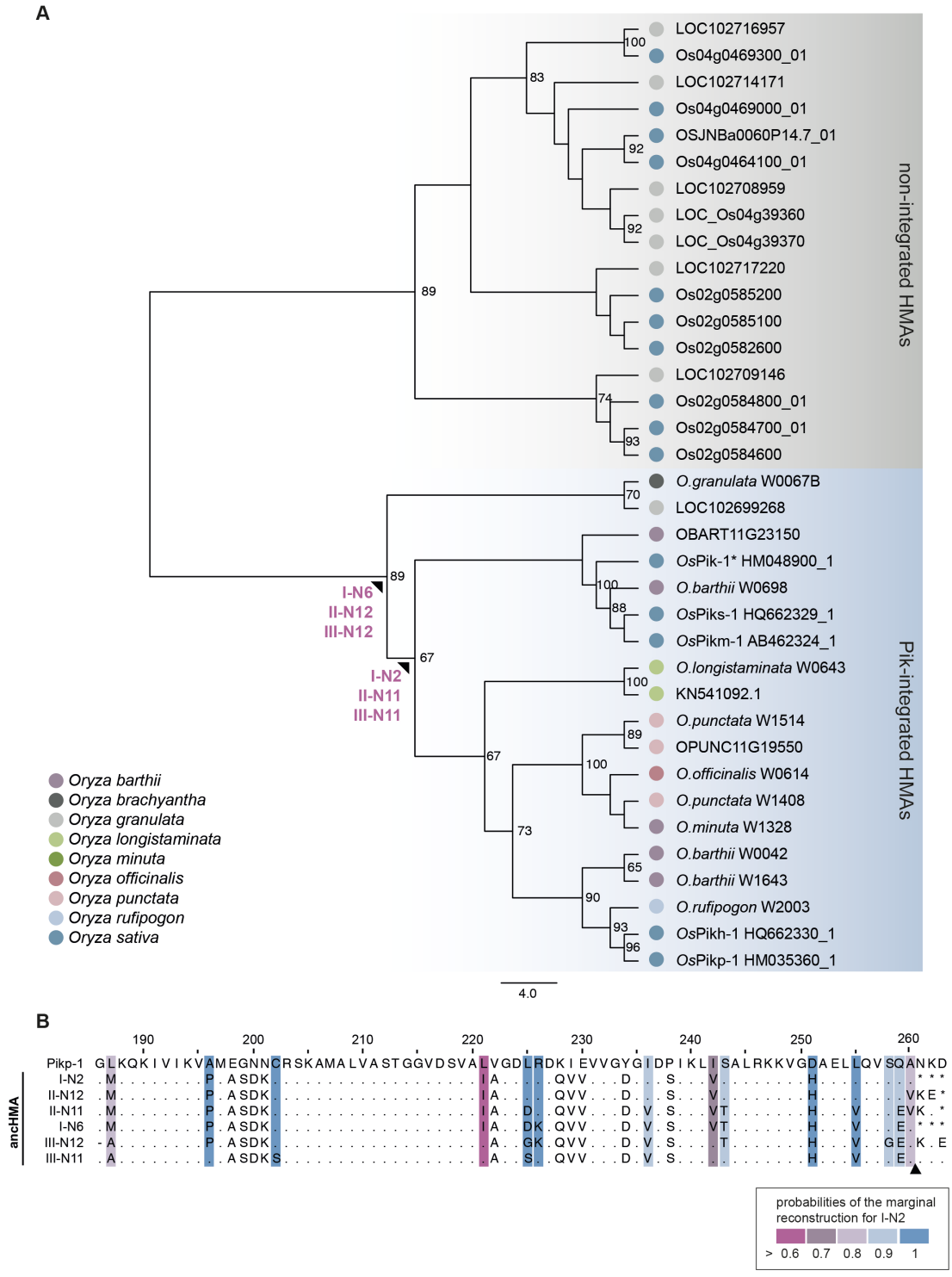


Figure 4.1 Ancestral sequence reconstruction yielded multiple plausible ancHMA sequences (continued)

(**A**) Representative neighbour joining phylogenetic tree of the HMA domain. The tree was built using JTT substitution model (Jones et al., 1992) and bootstrap method with 100 iterations (Felsenstein, 1985). Alignment of 98 amino acids of integrated (blue) and non-integrated (grey) HMAs was generated with MUSCLE (Edgar, 2004). Bootstrap values above 65% are shown at the base of respective clades. Nodes for which the ancestral sequence reconstruction was performed are marked with arrowheads. The scale bar indicates the evolutionary distance based on number of base substitutions per site. (**B**) Protein sequence alignment of representative ancestral HMA predictions. Amino acid sites for which sequence prediction was not performed are replaced with asterisks (*). The probabilities of the marginal reconstruction for I-N2 sequence are marked with coloured boxes. An arrowhead indicates the length of the construct used in further studies.

4.2.3 The LVKIE region of the Pikp-HMA domain determines AVR-PikD binding with high affinity

Next, I aimed to investigate which of the structural regions in the HMA encompasses adaptive mutations towards AVR-PikD binding. By synthesising sequence and structural information, I identified four polymorphic regions between the ancestral and modern Pikp-HMA (**Figure 4.2 A–B**). I sequentially replaced each of these regions in Pikp-1:ancHMA with the corresponding region from Pikp-HMA. Altogether, I obtained a suite of four chimeric HMAs—ancHMA$_{AMEGNND}$, ancHMA$_{LVKIE}$, ancHMA$_{LY}$, ancHMA$_{PI}$,—and assayed these for gain-of-binding to AVR-PikD in planta co-immunoprecipitation (co-IP) experiments. Among tested constructs only the Pikp-1:ancHMA$_{LVKIE}$ chimera associated with the effector at the levels that were similar to Pikp-1 (**Figure 4.2 C**; **Figure A.3.4**), indicating that residues in the LVKIE region are critical for enhanced effector binding.

Remarkably, the LVKIE region is located at the effector-binding interface, termed interface 2, where lysine (Lys-228) and glutamic acid (Glu-230) residues mediate the interaction between Pikp-HMA and AVR-PikD through hydrogen bonds or salt bridges (De la Concepcion et al., 2018; Maqbool et al., 2015). Amino acids at position 228 have also been assigned as a diagnostic marker for Pik breeding in rice (Costanzo and Jia, 2010). Interestingly, besides the LVKIE residues, ancHMA$_{LVKIE}$ is identical to ancHMA, which implies that the integrated HMA domain can accommodate mutations outside of the effector-binding interface without affecting the receptor's ability to bind the ligand.

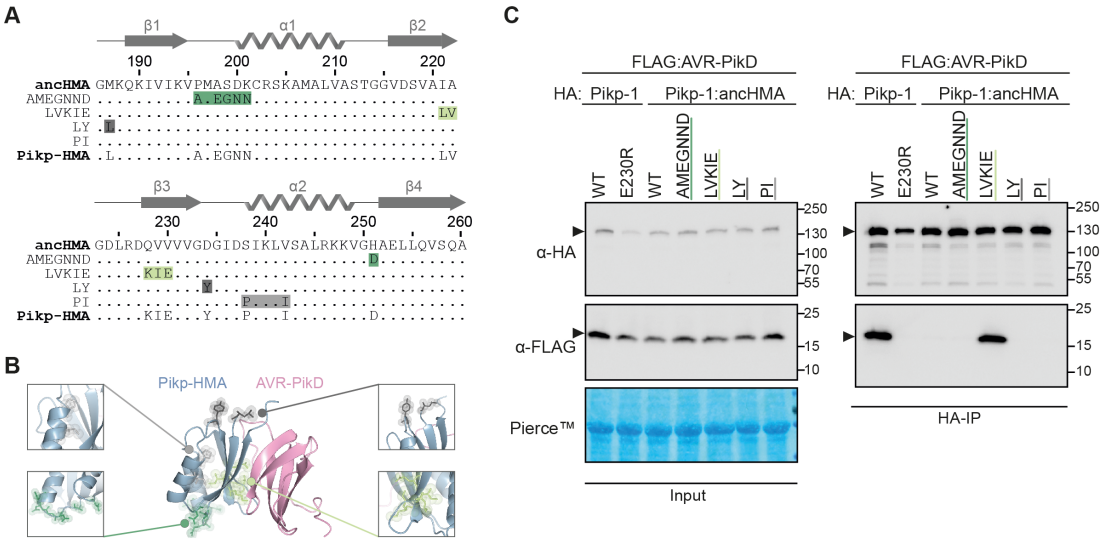


Figure 4.2 The integrated HMA domain of Pikp-1 evolved to interact with AVR-PikD with high affinity in the LVKIE region

(**A**) Protein sequence alignment showing the Pikp–ancHMA swap chimeras. The amino acid sequences of ancHMA, Pikp-HMA, and chimeras are aligned, with protein model above corresponding to the Pikp-HMA structure. The colour-coded rectangles correspond to polymorphic regions used for chimeric swaps. (**B**) Schematic representation of Pikp-HMA (blue) in complex with AVR-PikD (pink), with polymorphic regions between the Pikp-HMA and the ancHMA colour-coded as in the panel A. (**C**) The LVKIE substitutions in the ancestral HMA restore binding to AVR-PikD in the co-immunoprecipitation (co-IP) experiment. Association of FLAG-tagged AVR-PikD with HA-tagged Pikp-1, Pikp-1$_{E230R}$, Pikp-1:ancHMA, and Pikp-1:ancHMA chimeras, labelled above, was tested in in planta co-IP. Wild type (WT) Pikp-1 and Pikp-1$_{E230R}$ were used as positive and negative controls, respectively. Proteins obtained by co-immunoprecipitation with HA beads (HA-IP) and total protein extracts (Input) were immunoblotted with the appropriate antisera labelled on the left. Rubisco loading control was performed using Pierce™ staining. Arrowheads indicate expected band sizes. Results from three independent biological replicates of this experiment are shown in **Figure A.3.4**.

4.2.4 The Pikp-1 HMA domain only recently evolved to bind the AVR-PikD effector at high affinity

4.2.4.1 Two substitutions in the Pikp-HMA increased in planta association with AVR-PikD

To understand the mechanism by which LVKIE region facilitates the recognition of the AVR-PikD effector and the evolutionary trajectory of the binding, I set out to reconstruct the evolutionary history of this region. Probability-based ancestral sequence reconstruction, combined with hand-curation, was performed based on protein sequence

alignment and a representative phylogeny of 19 K-type integrated HMA domains, where ancHMA was separated from Pikp-HMA by five internal nodes (**Figure 4.1**). I identified three most ancient substitutions at the resolution of single amino acids—isoleucine 221 to leucin (Ile-221-Leu), followed by glutamine 228 to lysine (Gln-228-Lys), followed by valine 229 to isoleucine (Val-229-Ile) (**Figure 4.3 A**). Discerning the order of the two most recent substitutions, alanine 222 to valine (Ala-222-Val) and valine 230 to glutamic acid (Val-230-Glu), was not possible. I generated ancHMA mutants by consecutively introducing historical substitutions into their respective ancestral backgrounds, generating ancHMA$_{LAQVV}$, ancHMA$_{LAKVV}$, and ancHMA$_{LAKIV}$, as well as two plausible alternative states between LAKIV and LVKIE—ancHMA$_{LAKIE}$ and ancHMA$_{LVKIV}$. To determine the extent to which each of the historical mutations contributed to change in effector binding, I cloned the ancHMA mutants into Pikp-1 background and assayed them for AVR-PikD binding in planta. Initial results showed low accumulation level of Pikp-1:ancHMA$_{LVKIV}$ mutant, preventing meaningful interpretation of results obtained using this protein (**Figure A.3.5**), hence, I excluded it from further analysis; the remaining constructs accumulated to similar levels. In co-IP experiments, Pikp-1:ancHMA$_{LVKIE}$ exhibited the strongest association with AVR-PikD followed by Pikp-1:ancHMA$_{LAKIE}$, which displayed intermediate phenotype (**Figure 4.3 B**; **Figure A.3.6**). The remaining mutants did not show gain-of-binding to AVR-PikD when compared to ancHMA, indicating that the two most recent mutations, Ala-222-Val and Val-230-Glu, were instrumental in the HMA evolution towards AVR-PikD binding at high affinity.

4.2.4.2 Two substitutions in the Pikp-HMA increased in vitro association with AVR-PikD

To quantify how historical substitutions in the LVKIE region contributed to enhancement in binding to AVR-PikD, I carried out site-by-site surface plasmon resonance (SPR) experiment, using the AVR-PikD protein and the full set of the ancHMA mutants (cloned to match the residues Gly-186–Ser-258 of the full-length Pikp-1, which have previously been successfully used in vitro [Maqbool et al., 2015]). All proteins were purified from *E. coli* overexpression lines by two-step purification method (**Figure A.3.7**). Binding was measured by monitoring the relative response, following AVR-PikD immobilization on the NTA-sensor chip and injection of the ancHMA proteins at three

different concentrations. To capture binding dynamics, I recorded the response at two timepoints: at the end of HMA injection ('binding') and 15 seconds post-injection ('binding stability') (**Figure A.3.8 A**). I normalized the response units to the theoretical maximum response (R_{max}) and expressed the results as a percentage of R_{max} (%R_{max}), which gave a relative indication of binding strength. AncHMA$_{LVKIE}$ established the strongest interaction with AVR-PikD at levels similar to Pikp-HMA, followed by ancHMA$_{LAKIE}$, then ancHMA$_{LAQVV}$, ancHMA$_{LAKIV}$, and ancHMA, which showed weak interaction; I did not record any significant binding for ancHMA$_{LAKVV}$ (**Figure 4.3 C**; **Figure A.3.8 B**).

To estimate association/dissociation constant (K_D) for strong binders I monitored their interaction with AVR-PikD for extended concentration range (**Figure 4.3 D**; **Figure A.3.9 A–B**). The K_D values, calculated using one-to-one kinetics interaction model, ranged from 37.8 nM to 42.9 nM for ancHMA$_{LVKIE}$ (**Figure A.3.9 A–B**). AncHMA$_{LAKIE}$ bound to the effector with higher K_D, although this result was not successfully replicated due to insufficient data quality in one of the replicates. Despite lack of unspecific binding to the reference cell the positive control, Pikp-HMA, exhibited signal drop following the initial association phase, which precluded performing reliable K_D calculation based on kinetics interaction model. To circumvent these challenges, I calculated steady state affinity based on basic theory model, which can tolerate more data irregularities than kinetics-based models. Consistently with previous results, Pikp-HMA and ancHMA$_{LVKIE}$ showed comparable binding strength to AVR-PikD (114–157 nM and 147–159 nM, respectively), followed by slightly weaker affinity recorded for ancHMA$_{LAKIE}$ (423–493 nM) (**Figure 4.3 D**; **Figure A.3.9 C–D**). Due to insufficient data quality for some of the experiments and the overall error associated with the method the absolute K_D values should be interpreted with caution. Despite this, the overall patterns across the experiments were consistent, with Pikp-HMA and ancHMA$_{LVKIE}$ displaying the highest binding affinity towards AVR-PikD, followed by ancHMA$_{LAKIE}$. Altogether, I demonstrated that the two most recent substitutions enhanced HMA binding to the AVR-PikD effector in vitro.

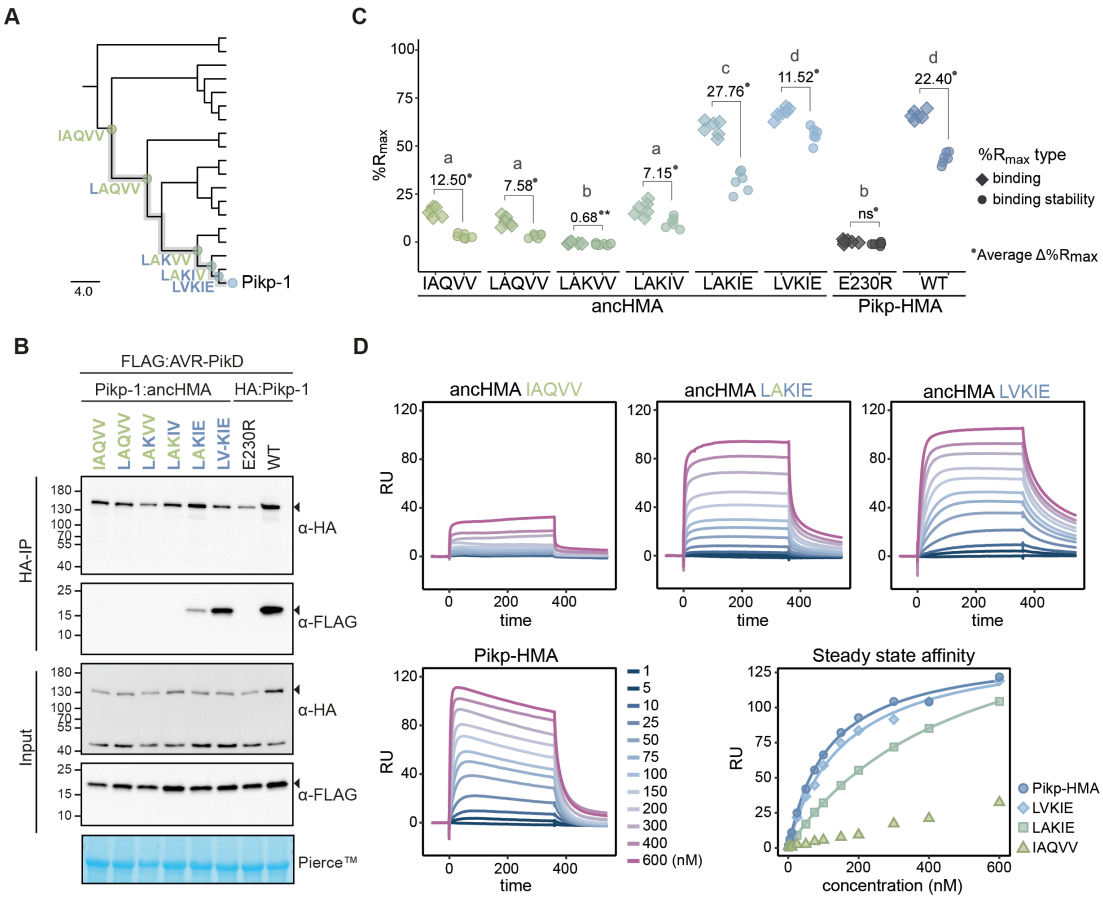


Figure 4.3 The Pikp-1 HMA domain only recently evolved the ability to bind the AVR-PikD effector at high affinity

(**A**) Schematic representation of the neighbour joining phylogenetic tree of the HMA domain from different *Oryza* species. The scale bar indicates the evolutionary distance based on number of base substitutions per site. Historical mutations in the LVKIE region acquired over the course of the Pikp-HMA evolution are shown next to the appropriate nodes. The mutations are colour-coded to match the ancestral (green) and present-day (blue) states. (**B**) Co-immunoprecipitation experiment illustrating in planta association of AVR-PikD with HA-tagged Pikp-1 and Pikp-1:ancHMA constructs, labelled above. The wild type (WT) HA:Pikp-1 and HA:Pikp-1$_{E230R}$ constructs were used as positive and negative controls, respectively. Proteins obtained by co-immunoprecipitation using HA resin (HA-IP) and total protein extracts (Input) were immunoblotted with the appropriate antisera labelled on the right. Loading control, featuring rubisco, was performed using Pierce™ staining. The arrowheads signify expected band sizes. The replicates of the experiment are shown in **Figure A.3.6**. (**C**) Plot illustrating calculated percentage of the theoretical maximum response (%R$_{max}$) values for interaction of HMA analytes, labelled below, with HIS:AVR-PikD ligand determined using surface plasmon resonance (SPR). %R$_{max}$ was normalized for the amount of ligand immobilized on the NTA-sensor chip. The chart summarises results obtained for HMA analytes at 400 nM concentration from three independent experiments with two internal repeats, with all the data points represented as diamonds or circles.»

Figure 4.3 The Pikp-1 HMA domain only recently evolved the ability to bind the AVR-PikD effector at high affinity (continued)

Three different concentrations of the analytes (400 nM, 200 nM, 50 nM) were tested; results for the 200 nM and 50 nM concentrations are shown in **Figure A.3.8 B**. Average $\Delta\%R_{max}$ (•) values represent absolute differences between values for 'binding' and 'biding stability' calculated from average values for each sample. Statistical differences among the samples were analysed with Tukey's honest significant difference (HSD) test ($p < 0.01$). P-values for all pairwise comparisons are presented in **Table A.3.2**. (**D**) The SPR multicycle kinetics sonsograms for association and dissociation of the AVR-PikD and HMA proteins. His-tagged AVR-PikD was immobilised on the sample cell, giving a response level of 157 ± 15 response units (RU). Steady state affinity biding curves were derived from multicycle kinetics sensograms. The replicates of the experiment and calculated K_D values are presented in **Figure A.3.9**.

4.2.4.3 Synthesis of the co-IP and SPR results

I was able to demonstrate that two substitutions in Pikp-HMA were critical for high affinity binding to AVR-PikD. In both co-IP and SPR experiments ancHMA displayed low affinity towards the effector, which was enhanced by Ala-222-Val and Val-230-Glu mutations.

In general, I observed a clear correlation between the co-IP and SPR experiments. Considering that co-IP experiments were performed using full-length proteins while SPR using only the HMA domains, this correlation implies that HMA alone is able to recapitulate effector binding preferences of the Pik-1 receptor. This is consistent with previous findings that have demonstrated that the HMA domains of Pikp-1 and Pikm-1 underly specificity and effector recognition profile (De la Concepcion et al., 2018; Maqbool et al., 2015). Notably, I observed differences in in planta and in vitro association strength between AVR-PikD and ancHMA$_{LAKIE}$. The difference between the two methods could be due to assay sensitivity, with SPR being able to detect small changes in binding. Alternatively, it's possible that other domains of Pik-1 can somehow modulate the binding, for example by restricting access to the HMA effector-binding interface. Similar discrepancies have previously been observed for Pikm-1–AVR-PikC interaction (De la Concepcion et al., 2020, 2019).

Taken together, these data demonstrate that for most of its evolutionary history Pikp-HMA had not been subject to selection pressure imposed by the AVR-PikD effector. I propose a model, in which the integrated HMA domain served as a bait for recognition of a different effector before it evolved new recognition specificity

towards AVR-PikD. Elevated mutation rates and strong signatures of selection precluded rigorous and accurate estimation of evolutionary divergence times of the integrated HMAs, however, there is wealth of evidence that indirectly suggests recent emergence of the LVKIE adaptations. Firstly, in the panel of 19 integrated HMA sequences collected in this study LVKIE is unique to only two of them, *Os*Pikp-1 and *Os*Pikh-1 (**Figure A.3.2**). Both Pikp-1 and Pikh-1 derived from domesticated rice, cloned from K60 japonica and Tetep indica cultivars, respectively (Jia et al., 2009; Wang et al., 2009). In addition, their genes are highly similar to each other; out of total of three polymorphisms, there is only one synonymous substitution that distinguishes their nearly 3,500-bp-long coding sequences. (For comparison, *Os*Pikp-1 differs from *Os*Pikm-1 from Tsuyake cultivar by 29 and 59 synonymous and non-synonymous substitutions, respectively.) All of the above provides indirect evidence that the Ala-222-Val and Val-230-Glu mutations indeed arose fairly recently, and it's interesting to speculate that their emergence coincided with the host jump of the blast fungus on domesticated rice.

4.2.5 Ala-222-Val and Val-230-Glu substitutions are sufficient to increase binding affinity towards the AVR-PikD effector

The role of historical contingency has been an important topic debated in the field of molecular evolution (Blount et al., 2018). Neutral mutations with no apparent effect on protein function might—and in fact very often do—fundamentally change protein evolutionary trajectory; some of them are thought to play a central role in evolution by permitting function-switching substitutions, hence the name permissive mutations (Harms and Thornton, 2013, 2010; Ortlund et al., 2007; Starr et al., 2017; Starr and Thornton, 2016). To investigate the role of contingency in the evolutionary history of the Pikp-1-integrated HMA domain, I tested the importance of early historical substitutions in the LVKIE region on effector binding strength. I incorporated Ala-222-Val and Val-230-Glu mutations into ancHMA, generating Pikp:ancHMA$_{IVQVE}$ construct, and examined effector binding in co-IP experiments (**Figure A.3.10**). Pikp:ancHMA$_{IVQVE}$ showed much stronger association with AVR-PikD than Pikp:ancHMA, however, I was unable to directly compare its association to Pikp:ancHMA$_{LVKIE}$ due to uneven protein accumulation levels.

These results demonstrate that Ala-222-Val and Val-230-Glu substitutions were sufficient to increase binding affinity towards the AVR-PikD effector, although one can't exclude the possibility that prior mutations had an additive effect on this interaction, not visible in co-IP. Another possibility, not addressed with this experiment, is that historical substitutions affected global protein structure rather than effector contact points. Permissive mutations have been shown to increase protein evolvability by increasing protein stability or folding (Harms and Thornton, 2010; Ortlund et al., 2007; Starr et al., 2017). It's interesting that Pikp:ancHMA$_{IVQVE}$ accumulation was reduced compared to Pikp:ancHMA$_{LVKIE}$, which may imply that gain-of-binding mutation had destabilising effect on the receptor and were historically contingent on the prior substitutions. Introducing gain-of-binding mutations into ancestral state of an entire Pikp-1 protein may shed light on the relevance of permissive mutations on Pikp-1 structure and function.

4.2.6 Structural determinants of binding

To understand the structural basis of gain-of-binding, I set out to determine co-crystal structures of the ancHMA and ancHMA$_{LVKIE}$ domains in complexes with AVR-PikD. Following heterologous protein overexpression in *E. coli* and two-step purification, I conducted a variety of commercially available crystallisation screens and optimisations. Obtained crystals were tested for X-ray diffraction at Diamond Light Source (Didcot, UK). Crystallisation screens resulted only in irregular crystals of ancHMA with AVR-PikD, which did not diffract under X-ray beam (data not shown). The best crystal of the ancHMA$_{LVKIE}$–AVR-PikD complex (purified and crystallised with assistance of Mauricio P Contreras, TSL) diffracted to 1.32 Å resolution. The structure was solved by molecular replacement using AVR-PikD structure in complex with Pikp-HMA as a template (De la Concepcion et al., 2018) and refined by iterative rounds of manual rebuilding and refinement with help from David Lawson from the John Innes Centre Crystallography Platform (Norwich, UK).

The ancHMA$_{LVKIE}$–AVR-PikD complex adopts a similar overall fold to previously published structures of Pikp-HMA and Pikm-HMA in complexes with the AVR-Pik effectors (**Figure 4.4 A**; **Table A.3.3**). The complex constitutes a single effector molecule, positioned as in other published structures, and two HMA molecules; HMA dimerization is most likely an artefact of protein expression and in vitro purification and

is not biologically relevant (Maqbool et al., 2015). The core HMA structure is composed of two α-helices, involved in HMA homodimerization, and four antiparallel β-strands, located on the opposite site, which form an extensive effector-binding interface. The interface buries 19% of total solvent-accessible surface area of the HMA and is dominated by hydrogen bonds formed between the two peptides or water molecules coordinated at the interface. Direct interaction between the HMA_{LVKIE} and AVR-PikD is mediated by nine residues on each side, with the main contributions derived from Ser-218, Glu-230, Lys-228, and Asp-224. Side chain hydroxyl groups of serine 218 and glutamic acid 230 form hydrogen bonds or salt bridges with AVR-$PikD^{His-46}$; lysine 228 forms a salt bridge with AVR-$PikD^{Asp-66}$ through side chain nitrogen; aspartic acid 224 forms two salt bridges/hydrogen bonds with AVR-$PikD^{Arg-64}$.

Failed crystallisation trials for the ancHMA–AVR-PikD complex precluded direct comparison of the LVKIE region between ancHMA and $ancHMA_{LVKIE}$. To gain insights into the structural determinants of effector binding in this region, I performed homology modelling and generated a model of the ancHMA in complex with AVR-PikD. I further validated modelled interactions by examining a published structure of Pikm-HMA (De la Concepcion et al., 2018), whose interface 2 is identical to the interface 2 present in ancHMA. As expected, ancHMA was predicted to adopt a similar fold and interact with the effector through the same interface as its $ancHMA_{LVKIE}$ derivative (**Figure A.3.11**). Close inspection of the structures revealed that, in contrast to Glu-230 of $ancHMA_{LVKIE}$, much smaller valine located at the structurally equivalent position in ancHMA and Pikm-HMA fails to form a hydrogen bond with AVR-$PikD^{His-46}$ (**Figure 4.4 B**). Although this histidine is still coordinated by conserved Ser-218, loss of hydrogen bond at position 230 most likely results in substantial reduction in ancHMA–AVR-PikD association strength, as observed in in vitro and in planta binding experiments (**Figure 4.3**). The impact of other mutations in the LVKIE region is not immediately apparent. This is in line with results from the SPR experiments, which showed that those substitutions have no or only a marginal effect on binding in vitro. The Ala-222-Val mutation did, however, cause a clear enhancement of association with the effector in co-IP experiment. Why? One may speculate that it might increase binding robustness and improve the interaction in the context of the full-length NLR.

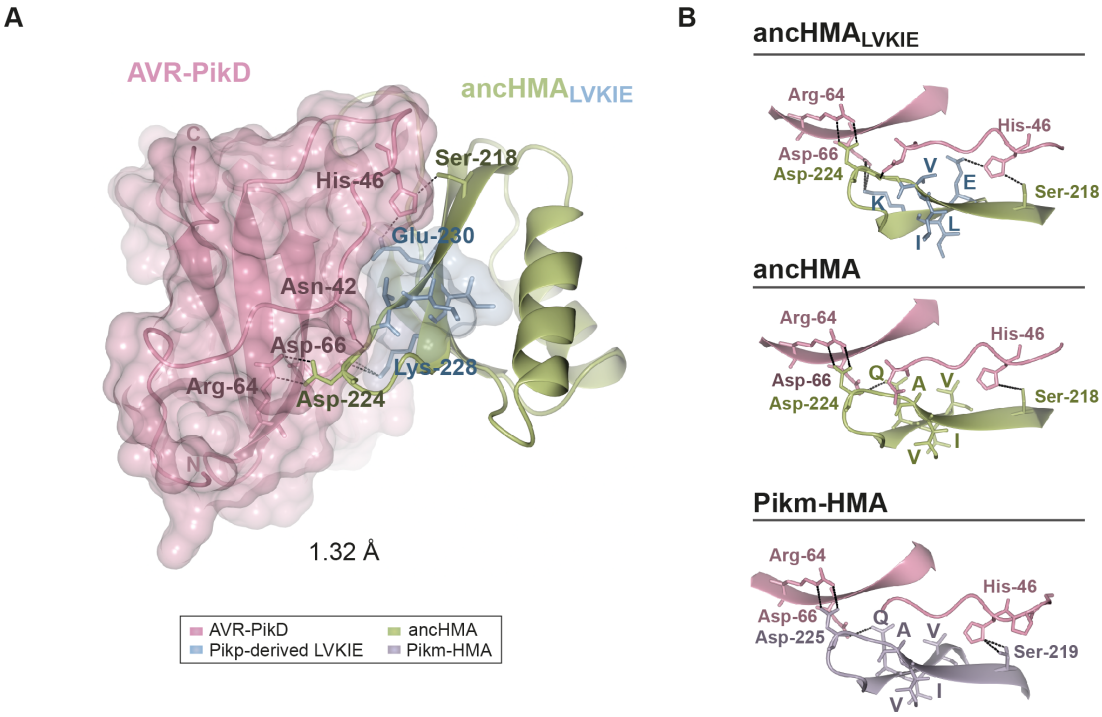


Figure 4.4 Val-230-Glu mutation within the LVKIE region of the ancHMA enhances interaction with AVR-PikD through hydrogen bond formation

(**A**) Schematic representation of the structure of ancHMA$_{LVKIE}$ with the AVR-PikD effector. The molecules are shown as ribbons with selected side chains presented as sticks and labelled; the colours of the residue labels match colours of the respective molecules. The molecular surfaces of the AVR-PikD (pink) and LVKIE residues (blue) within ancHMA$_{LVKIE}$ are also shown. Dashed lines represent hydrogen or disulfate bonds formed between the two molecules. (**B**) Close-up views of the interaction interface 2 of ancHMA$_{LVKIE}$ (green and blue), ancHMA (green), and Pikm-HMA (purple) showcasing differences in binding to AVR-PikD (pink). The selected residues involved in binding are labelled with labels matching the colours of the corresponding molecules. The LVKIE residues are labelled with single-letter amino acid symbols—Ile/Leu-221 (I/L), Ala/Val-222 (A/V), Gln/Lys-228 (Q/K), Val/Ile-229 (V/I), Val/Glu-230 (V/E) for ancHMA/ancHMA$_{LVKIE}$; and Ile-222 (I), Ala-223 (A), Gln-231 (Q), Val-232 (V), Val-233 (V) for Pikm-HMA. Dashed lines represent hydrogen or disulfate bonds formed between the two molecules.

4.2.7 High binding affinity to AVR-PikD accounts for the ability to trigger an immune response

4.2.7.1 The Pikp-1:ancHMA fusions are autoactive in Pikp-2–dependent manner

To test if effector binding by Pikp-1:ancHMA is sufficient to trigger an immune response, I performed hypersensitive response (HR) experiment by transiently

co-expressing each of the Pikp-1:ancHMA fusions with AVR-PikD and Pikp-2 in *Nicotiana benthamiana*. I discovered that all Pikp-1:ancHMA variants were autoactive and triggered spontaneous cell death in the absence of the effector (**Figure 4.5**; **Figure A.3.12**; **Figure A.3.13**). Notably, the presence of Pikp-2 partner was required for autoactivity, which is consistent with previously published results (Zdrzałek et al., 2020).

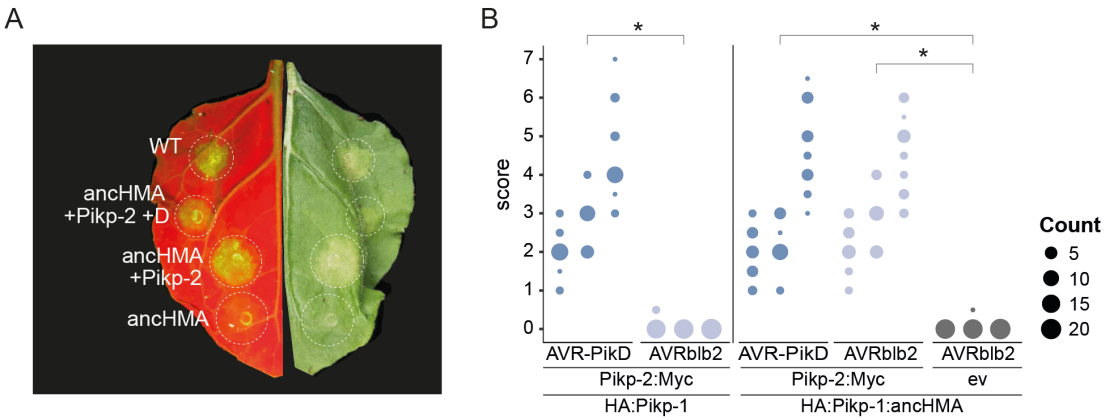


Figure 4.5 The Pikp-1:ancHMA fusion is autoactive in Pikp-2–dependent manner

Wild type (WT) Pikp-1 and Pikp-1:ancHMA (ancHMA) were transiently co-expressed with AVR-PikD (D)/AVRblb2 and Pikp-2/empty vector (ev). AVRblb2, an effector from the potato blight pathogen *Phytophthora infestans*, and the empty vector were used as negative controls. (**A**) Photos of representative *Nicotiana benthamiana* leaves infiltrated with appropriate constructs (labelled on the left) photographed five days after infiltration under UV (left) and day light (right). (**B**) Hypersensitive response (HR) was scored at five days post-agroinfiltration. The results are presented as a dot plot, where a size of a dot is proportional to the number of samples with the same score (count) within the same biological replicate. The experiment was independently repeated at least three times with 22–24 internal replicates; the columns within tested conditions (labelled on the bottom) correspond to results from different biological replicates. Significant differences between the conditions are indicated with asterisk (*). The details of statistical analysis are presented in **Figure A.3.13**. The experiment was performed with different variants of plausible ancHMA sequences summarised in **Figure A.3.12**.

4.2.7.2 Amino acid substitutions in the β1–α1 and α2–β4 loops of the Pikp-HMA domain abolish the autoactivity phenotype

Next, I used previously generated fusions with chimeras of the ancHMA—ancHMA$_{AMEGNND}$, ancHMA$_{LVKIE}$, ancHMA$_{PI}$, ancHMA$_{LY}$—as a proxy to delimitate the region responsible for the autoactivity phenotype of Pikp-1:ancHMA. I tested these fusions for loss-of-function in HR assay by transient co-expression with Pikp-2

in *Nicotiana benthamiana* (**Figure 4.6**; **Figure A.3.14; Figure A.3.15**). Among these, Pikp-1:ancHMA$_{AMEGNND}$, but no other mutant, showed complete loss-of-autoactivity. This phenotype was not due to protein instability or low protein abundance (**Figure A.3.4**), suggesting that the AMEGNND region, located in the β1–α1 and α2–β4 loops of the Pikp-HMA domain, underpins Pikp-1:ancHMA autoactivity.

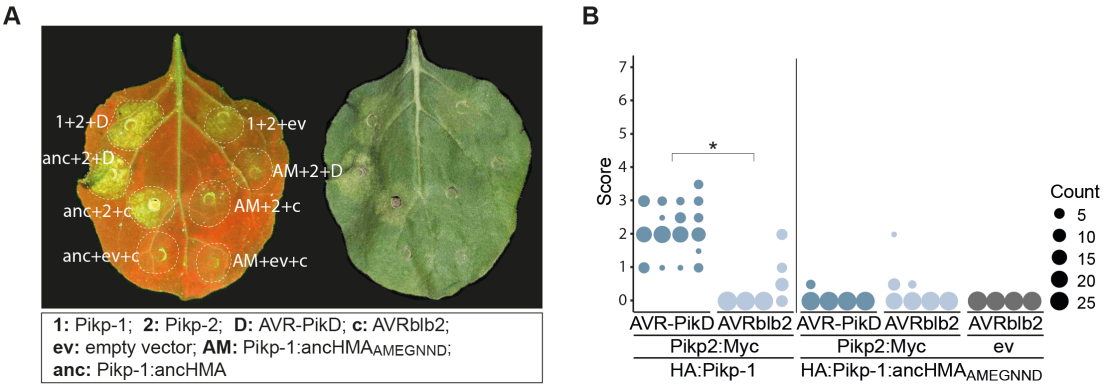


Figure 4.6 Mutations within the AMEGNND region abolish ancHMA-triggered autoactivity

Cell death assay after transient co-expression of Pikp-1:ancHMA$_{AMEGNND}$ with AVR-PikD/ AVRblb2 and Pikp-2/ev. AVRblb2 and the empty vector (ev) were used as negative controls. (**A**) Representative *Nicotiana benthamiana* leaves infiltrated with appropriate constructs (labelled next to the infiltration spot) were photographed five days after infiltration under UV (left) and day light (right). (**B**) Hypersensitive response (HR) was scored five days after agroinfiltration. The results are presented as a dot plot where the size of a dot is proportional to the number of samples with the same score (count) within the same biological replicate. The experiment was independently repeated at least four times with 24–26 internal replicates; the dot columns within tested conditions (labelled on the bottom) present results from different replicates. Significant differences between the conditions are shown with asterisk (*). The details of statistical analysis are summarised in **Figure A.3.15**.

4.2.7.3 Strong binding to the effector results in Pik-mediated cell death

To investigate the link between AVR-PikD binding and immune response, I performed HR assay in *N. benthamiana* leaves using Pikp-1:ancHMA mutants in the LVKIE region. Leveraging the information about autoactivity determinants, I first removed autoactivity by introducing AMEGNND mutations into these constructs (**Figure 4.7 A**), henceforth called Pikp-1:ancHMA$_{LVKIE}$*, Pikp-1:ancHMA$_{LAKIE}$*, Pikp-1:ancHMA$_{LAKVV}$*, Pikp-1:ancHMA$_{LAQVV}$*. None of resulting mutants triggered

spontaneous cell death when transiently co-expressed with Pikp-2 (**Figure 4.7 B**). Co-expression with AVR-PikD revealed that the strength of binding directly correlated with the strength of HR: Pikp-1:ancHMA*, Pikp-1:ancHMA$_{LAKVV}$*, and Pikp-1:ancHMA$_{LAQVV}$* failed to elicit immune signalling; Pikp-1:ancHMA$_{LVKIE}$* showed strong cell death phenotype; whereas the response triggered by the intermediate binder Pikp-1:ancHMA$_{LAKIE}$* was slightly, yet significantly, reduced when compared to Pikp-1:ancHMA$_{LVKIE}$* (**Figure 4.7**;**Figure A.3.16**). All proteins accumulated at similar levels in western blot analysis (**Figure A.3.17**). Overall, these results indicate that adaptations of the HMA domain towards strong effector binding enabled effector-dependent activation of immune signalling and robust immune response.

The above results are in agreement with previous studies showing that only high affinity interaction with the effector accounts for Pikp-mediated immune response (De la Concepcion et al., 2020, 2019, 2018; Maqbool et al., 2015). Furthermore, there is an indication that the strength of effector binding may have a quantitative effect on the strength of immune response, as illustrated by Pikp-1:ancHMA$_{LAKIE}$* and Pikp-1:ancHMA$_{LVKIE}$*. Visual scoring of the intensity of HR has been shown to capture the quantitative differences in cell death severity (De la Concepcion et al., 2019), still, it remains a semi-quantitative method and more accurate measurements would be necessary to confirm observed differences.

The mechanism underlying NLR activation upon effector binding and the extent to which the binding affinity contributes to immune response have been a subject of debate (Bialas et al., 2018). Binding above a certain threshold could be important either to guarantee the right amount of effector-bound NLR molecules to trigger immune response or for inducing conformational changes necessary for NLR activation. It is important to note, though, that overexpression of NLR and effector proteins in a model plant such as *N. benthamiana* may not fully reflect rice–rice blast interaction in nature. Future studies might bridge this gap and reveal if the differences in the strength of immune response affect Pikp-mediated resistance. Would Pikp-1:ancHMA$_{LAKIE}$* mediate resistance towards *Magnaporthe oryzae* carrying AVR-PikD? If so, would the resistance be weaker? Taken together, these data suggest that Ala-222-Val and Val-230-Glu mutations derived as a result of Pikp-HMA adaptation to the AVR-PikD effector and were critical for effector recognition and subsequent Pikp-1–mediated immune response.

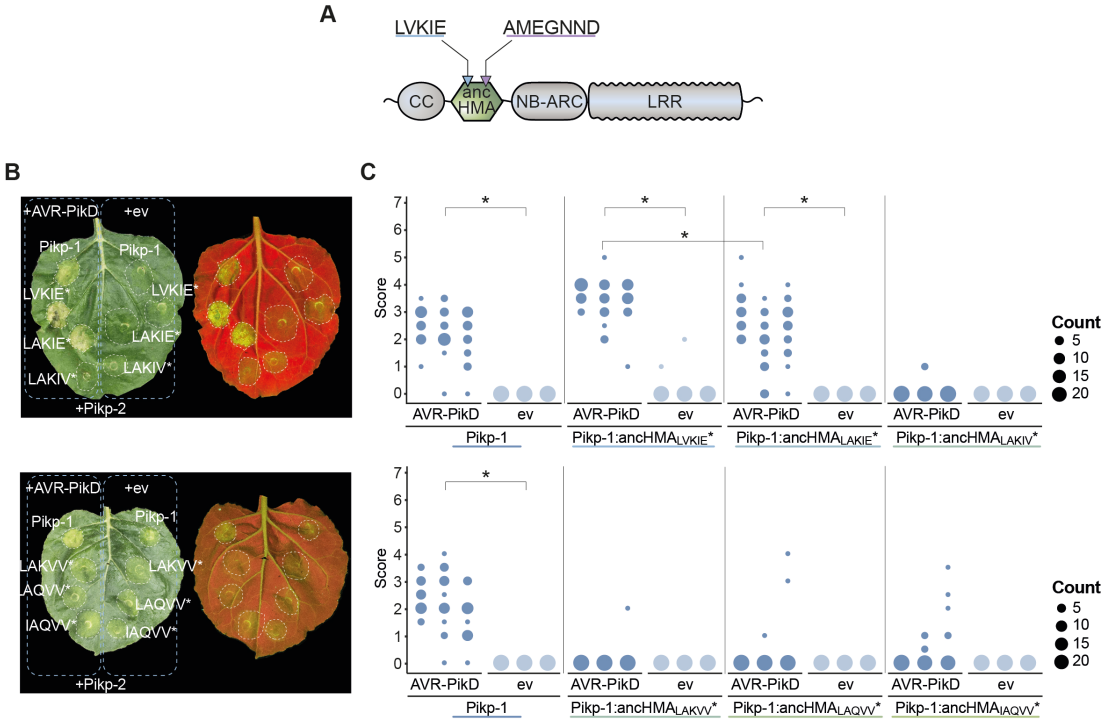


Figure 4.7 Pikp-1:ancHMA$_{LVKIE}$* and Pikp-1:ancHMA$_{LAKIE}$* mediate immune response to the AVR-PikD effector

(**A**) Schematic representation of the strategy used for removing autoavtivity from the Pikp-1:ancHMA fusions. The mutated regions are presented with arrowheads. (**B**) Representative images of cell death assay after transient co-expression of the Pikp-1:ancHMA* mutants, featuring HF tag—Pikp-1:ancHMA$_{LVKIE}$*, Pikp-1:ancHMA$_{LAKIE}$*, Pikp-1:ancHMA$_{LAKIV}$*, Pikp-1:ancHMA$_{LAKVV}$*, Pikp-1:ancHMA$_{LAQVV}$*, Pikp-1:ancHMA$_{IAQVV}$*—with AVR-PikD. Empty vector (ev) was used as a negative control. All constructs were co-expressed with the gene silencing suppressor p19 (Win and Kamoun, 2003). The leaves were photographed five days after infiltration under day light (left) and UV light (right). (**C**) Hypersensitive response (HR) was scored at five days post-agroinfiltration. The results are presented as dot plots, where a size of a dot is proportional to the number of samples with the same score (count) within the same biological replicate. The experiment was independently repeated at least three times with 23–24 internal replicates; the columns within tested conditions (labelled on the bottom) correspond to results from different biological replicates. Significant differences between selected conditions are marked with asterisk (*). The details of statistical analysis are summarised in **Figure A.3.17**.

4.3 Conclusions

In conclusion, both in planta and in vitro experiments indicated that the integrated HMA domain of Pikp-1 has evolved a higher binding affinity to the AVR-PikD effector from a weaker ancestral state. Not only did I demonstrate that the mutations at the effector-binding interface 2 have been essential for increasing the strength

of the interaction between HMA and AVR-PikD, but I was also able to pinpoint how individual mutations acquired by the Pikp-HMA domain contributed to change in effector binding and the strength of subsequent immune response. These results revealed that the Pikp-HMA domain only recently acquired the capacity to bind AVR-PikD at high affinity. In fact, it appears that for most of its evolutionary history the HMA domain had not been subject to selection pressure to recognise the AVR-PikD effector. I propose that the integrated HMA domain recognized different pathogen effector(s) for millions of years before recently 'switching' to AVR-PikD, possibly following the blast jump to domesticated rice. Furthermore, this work demonstrates that the synthesis of mechanistic and evolutionary research can provide a robust framework to study NLR-IDs. Using this line of research, future studies can address a number of unanswered questions. These include, but are not limited to, the coevolution of the integrated HMA with other domains of Pikp complex and a potential role that other blast effectors might have played in the evolution of Pikp-1.

Chapter 5: AVR-PikD recognition by Pikm-1 evolved through a different path than in Pikp-1—structural basis of convergent molecular evolution

5.1 Introduction

Frequent changes in effector repertoires of plant pathogens shape the plant immune system, leading to the emergence of diverse immune receptors that provide robust and broad-spectrum disease resistance. As a consequence, NLR repertoires and diversification patterns within NLR genes, are thought to reflect local adaptations to pathogen populations (Liao et al., 2016). For instance, the *Pm3* gene mediating resistance to powdery mildew displays high interspecies diversity between *Triticum aestivum* and *T. dicoccoides* whose distinct evolutionary histories have been proposed to result from different interactions with specialised pathogen lineages (Sela et al., 2014). Similarly, the alleles of *L* locus of flax display differences in recognition specificity of the AvrL567 effector alleles from flax rust fungus (Dodds et al., 2006), with some *L* alleles mediating resistance against unrelated effectors (Dodds et al., 2004).

Extreme patterns of natural selection can lead to independent emergence of similar phenotypic features in distinct plant populations—a phenomenon known as convergent evolution. Convergent phenotypes can be produced through similar changes at the genetic level, or may have different genetic basis (Blount et al., 2018; Losos, 2011; Rokas and Carroll, 2008; Washburn et al., 2016). Plant–microbe systems are exceptional in the number of examples of repeated evolution of similar traits, many of which are likely a result of pathogen-imposed selection (Tamborski and Krasileva, 2020; Upson et al., 2018). One example comes from phylogenetically unrelated NLR proteins that converged on similar recognition mechanisms or recognition specificities, like Arabidopsis RPM1 and RPS2, and RPG1-b from soybean, all of which recognise RIN4 perturbations induced by the AvrB effector from *P. syringae* (Ashfield et al., 2004; Kessens et al., 2014; Selote and Kachroo, 2010). Despite these cases, the molecular mechanisms underlying convergent evolution of effector recognition in NLRs are largely unknown.

As discussed in previous chapters (**Chapters 1**, and **4**), the Pik-1 receptor mediates resistance to the *M. oryzae* fungus by direct binding of the AVR-Pik effector via its

integrated HMA domain (Kanzaki et al., 2012; Maqbool et al., 2015). Both Pik and AVR-Pik occur in allelic series, with the Pik proteins showing different recognition specificities against AVR-Pik variants (De la Concepcion et al., 2020, 2018; Kanzaki et al., 2012; Li et al., 2019; Longya et al., 2019). Studies dissecting the allelic diversity of Pik revealed that the Pikp allele, from K60 rice cultivar (Wang et al., 2009), can recognise only the AVR-PikD effector allele, whereas the Pikm allele, from the Tsuyuake cultivar (Ashikawa et al., 2008), shows much broader recognition spectrum, recognising not only AVR-PikD but also AVR-PikA and AVR-PikE alleles (De la Concepcion et al., 2018; Kanzaki et al., 2012). These recognition specificities have been linked to effector binding affinity (De la Concepcion et al., 2018; Maqbool et al., 2015), and are thought to reflect the ongoing arms race between rice and the rice blast fungus (Białas et al., 2018; Kanzaki et al., 2012; Li et al., 2019). This is consistent with the extreme rates of protein diversification observed within the Pik-1–integrated HMA domain (**Chapter 3**). Interestingly, as I showed earlier, the Pikp-1 and Pikm-1 fall into two phylogenetically distinct clades of Pik-1 (**Chapters 3** and **4**) (De la Concepcion et al., 2020; Kanzaki et al., 2012) and appear to have evolved their effector-recognition specificities independently.

The evolutionary trajectory of the Pikp-HMA domain (**Chapter 4**) elegantly illustrates how changes in the environment, such as new pathogen threats, can lead to rapid adaptations within plant immune receptors. However, sequence comparisons and the aforementioned phylogenetic analyses indicated that similar phenotypic adaptations of Pik alleles—namely, recognition of the AVR-PikD effector—seem to have arisen as a result of different mutational pathways, reminiscent of convergent evolution. In this chapter, I aimed to test this hypothesis and understand the alternative evolutionary pathways that led to the emergence of different Pik alleles of rice that recognise the same blast effectors. Here, I leveraged the experimental framework established for Pikp-HMA (**Chapter 4**) and reconstructed the evolutionary trajectory of Pikm-HMA from their last common ancestor, the ancHMA. Using in vitro and in planta biochemistry methods, I functionally characterised a region in the Pikm-HMA—different from the one identified for Pikp-HMA—that evolved towards AVR-PikD binding at high affinity. Together, these experiments revealed that Pikp and Pikm underwent different evolutionary pathways to produce similar phenotypic outcome, painting a complex picture of the dynamics of their evolution at the population level.

5.2 Results and discussion

5.2.1 The EMVKE region of the integrated HMA of Pikm-1 determines AVR-PikD binding with high affinity

In the previous chapter (**Chapter 4**), I identified the LVKIE region that encompasses Pikp-1 adaptive mutations towards AVR-PikD binding and recognition. Notably, these adaptations are rare among Pik-1 alleles and orthologues in *Oryza* (**Figure A.3.2**), and in most cases this sequence has remained largely unchanged from the predicted ancestral IAQVV state. Pikm-1, a Pik-1 allele present in the Tsuyuake rice cultivar (Ashikawa et al., 2008), is an example of a Pik-1 receptor that carries IAQVV residues. Despite the absence of the adaptive LVKIE mutations, Pikm-1 binds the AVR-PikD effector with high affinity and triggers immune response upon effector recognition (De la Concepcion et al., 2018; Kanzaki et al., 2012). This led to the hypothesis that the integrated HMA domain of Pikm-1 (Pikm-HMA) has undergone an evolutionary pathway towards effector recognition different from that of Pikp-HMA, and that its effector-driven adaptative mutations lay outside of the LVKIE region.

Which Pikm-HMA mutations have enabled AVR-PikD binding? To address this question, I performed structure-informed sequence comparison of the Pikm-HMA and ancHMA domains. First, to ensure that the full diversity within the HMA domain was captured, I amended the ancHMA sequence with a three-amino-acid-long extension, corresponding to residues 262–264 of the full-length Pikm-1. Although these three residues are polymorphic in Pikm-HMA they are identical between ancHMA and Pikp-HMA and were thus previously omitted in the studies concerning Pikp-HMA (**Chapter 4**). Next, using sequence and structural information, I identified five polymorphic regions differentiating the ancestral HMA and modern Pikm-HMA (**Figure 5.1 A–B**). To determine, which of those regions underpin AVR-PikD binding, I first introduced ancHMA into Pikm-1 to create an NLR sensor that mimics the ancestral state. Next, I sequentially incorporated present-day residues from each of these regions into the Pikm-1:ancHMA fusion, and tested them for interaction with the effector using co-immunoprecipitation (co-IP). Among five chimeras tested in this experiment, only Pikm-1:ancHMA$_{EMVKE}$ associated with AVR-PikD (**Figure 5.1 C**; **Figure A.4.1**). I failed to detect any interaction with the effector for Pikm-1:ancHMA and the remaining chimeras, whereas Pikm-1:ancHMA$_{VH}$ was unstable, hence results obtained using this

construct were inconclusive. Overall, these results suggest that Pikm-HMA evolved towards AVR-PikD recognition through mutations in the EMVKE region that determines high affinity binding to this effector.

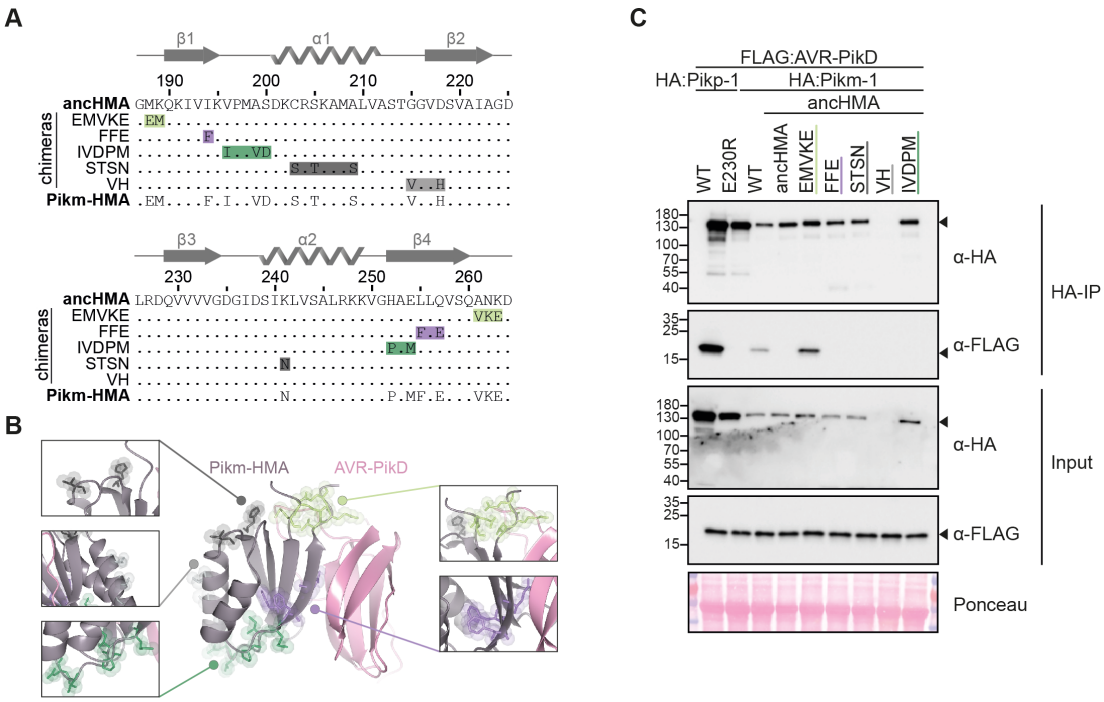


Figure 5.1 Adaptive mutations towards AVR-PikD binding are located in the EMVKE region of the HMA domain of Pikm-1

(**A**) Protein sequence alignment between ancHMA, Pikm-HMA, and Pikm:ancHMA chimeras. Protein model above the alignment depict Pikm-HMA structure. The colour-coded rectangles mark polymorphic regions used for chimeric swaps. (**B**) Schematic representation of the Pikm-HMA domain (purple) in complex with AVR-PikD (pink), with polymorphic regions between Pikm-HMA and ancHMA are colour-coded as in panel A. (**C**) EMVKE substitutions in the ancestral HMA restore in planta association with AVR-PikD. Co-immunoprecipitation experiment of FLAG-tagged AVR-PikD with the ancHMA chimeras, featuring HA tag, labeled above. Wild type (WT) Pikp-1/Pikm-1 and Pikp-1_{E230R} were used as positive and negative controls, respectively. Proteins obtained by co-immunoprecipitation with HA beads (HA-IP) and total protein extracts (Input) were immunoblotted with the appropriate antisera labelled on the right. Rubisco loading control was carried out using Ponceau staining solution. Arrowheads, shown on the right, indicate expected band sizes. The replicates of the co-IP experiment are shown in **Figure A.4.1**.

Three predominant regions of the HMA–effector interaction surface, termed interfaces 1, 2, and 3, have been described for Pik-1 receptors (De la Concepcion et al., 2018). The EMVKE residues are part of the interfaces 1 and 3, located at the β-strand 1 (β1) and β-strand 4 (β4), respectively. Residues at the interface 3 of Pikm-HMA have been shown to mediate tight interaction with the AVR-PikD effector—supposedly to compensate for weak interaction at the interface 2—and to underpin broad effector-recognition specificity (De la Concepcion et al., 2020, 2019, 2018). Notably, residues located at all three interfaces show relatively high levels of nonsynonymous sequence substitutions in comparison to synonymous mutations, with four out of five sites from the EMVKE region exhibiting moderate or high posterior probability of positive selection (**Chapter 3**).

5.2.2 Mutations located at the binding interface 3 enabled Pikm-1 evolution towards AVR-PikD binding

5.2.2.1 Reconstruction of the Pikm-HMA evolutionary pathway towards AVR-PikD binding

In order to trace the evolutionary trajectory of the Pikm-HMA domain that resulted in AVR-PikD binding, I reconstructed the mutational history of the EMVKE effector-binding interface (**Figure 5.2 A**). The ancestral sequence reconstruction was performed by combination of manual and probability-based approaches using an amino acid sequence alignment and a representative phylogenetic tree of the HMA domain, where Pikm-HMA and ancHMA were separated by four internal nodes (**Figure 4.1**). According to the ancestral sequence prediction, the ancestral state of the HMA domain carried MKANK residues in the EMVKE region. I was able to identify one node that represents an evolutionary intermediate between the ancestral MKANK and present-day EMVKE states, namely EMANK, that emerged through MK>EM mutations—methionine 188 to glutamate (Met-188-Glu) and lysine 189 to methionine (Lys-189-Met). The ANK>VKE mutations—alanine 261 to valine (Ala-261-Val), asparagine 262 to lysine (Asp-262-Lys), and lysine 263 to glutamate (Lys-263-Glu)—were acquired at the later timepoint. Given the resolution of the phylogenetic tree, determining the order of individual mutations was not possible.

To evaluate the effect of these historical mutations on effector binding, I generated the ancHMA mutant that recapitulated the predicted step-by-step intermediate state of the EMVKE region, ancHMA$_{EMANK}$, incorporated this mutant into the HA:Pikm-1 backbone, and tested its in planta association with AVR-PikD (**Figure 5.2 B**; **Figure A.4.2**). Pikm:ancHMA$_{EMVKE}$ exhibited strong association with AVR-PikD in co-IP experiments, consistent with my previous results (**Section 5.2.1**). In contrast, Pikm:ancHMA$_{EMANK}$ did not show gain-of-binding relative to Pikm:ancHMA$_{MKANK}$. This suggests that ANK>VKE substitutions were essential for the Pikm-HMA evolution towards AVR-PikD binding at hight affinity.

5.2.2.2 In vitro validation of differential binding to AVR-PikD across the Pikm-HMA evolutionary states

Next, I aimed to validate the binding of the step-by-step evolutionary transition of the EMVKE region in vitro, using the AVR-PikD protein and the full set of the ancHMA mutants purified from *E. coli* overexpression lines (**Figure A.4.3**). To encompass full diversity between the ancestral and present-day states of the Pikm-HMA, I used HMA sequences that consisted of a five–amino acid extension at the C-terminus compared to the constructs used in studies concerning the evolutionary trajectory of Pikp-HMA (**Chapter 4**), herein called ancHMA+5. During the purification process I noted a shift in elution volume of the ancHMA+5 in complex with AVR-PikD relative to the elution volume of the ancHMA$_{LVKIE}$–AVR-PikD complex in size-exclusion chromatography (**Figure A.4.4**). I inferred that this difference is a result of different stoichiometries of the ancHMA–AVR-PikD complexes; while ancHMA$_{LVKIE}$–AVR-PikD formed two-to-one complex (**Chapter 4**), the constructs with the extension interacted with the effector at one-to-one ratio.

To validate and quantify the binding of the step-by-step evolutionary transition of the EMVKE region in the HMA domain, I carried out surface plasmon resonance (SPR) experiments using the purified proteins (**Figure A.4.3**). I measured the binding by flowing the HMA analytes over the AVR-PikD protein immobilised on the NTA-chip and monitoring relative response (expressed in response units [RU]). To capture binding dynamics, I recorded the response at two timepoints: at the end of analyte injection ('binding') and 15 seconds post-injection ('binding stability'). I calculated the binding

strength as a percentage of R_{max} (%R_{max}), assuming one-to-one interaction, with the results normalised against theoretical maximum response (R_{max}), which gives relative indication of binding affinity. Surprisingly, at 400 nM concentration all tested HMA variants bound to the AVR-PikD effector with similar strength (**Figure 5.2 C**; **Figure A.4.5**; **Table A.4.1**). However, there were marked differences in the association and dissociation rates as well as overall binding dynamics between tested constructs. First, despite high binding values, ancHMA exhibited fast dissociation rates as illustrated by the pattern of 'binding stability' (**Figure 5.2 C**; **Figure A.4.5**; **Table A.4.2**) and shape of the curves (**Figure 5.2 D**; **Figure A.4.6**). Second, ancHMA$_{EMVKE}$ displayed high values for 'binding' and 'binding stability', with a gentle slope for association across the concentrations, indicating tight and stable binding. Finally, ancHMA$_{EMANK}$ fell in-between ancHMA and ancHMA$_{EMANK}$, with stable and relatively strong binding at the top concentration and moderate binding strength at lower concentrations. The above indicates that ANK>VKE substitutions were essential for the Pikm-HMA evolution towards AVR-PikD binding at hight affinity.

These results could have been validated by calculating association/dissociation constant (K_D) for HMA–AVR-PikD interaction; unfortunately, due to insufficient data quality obtained from multicycle kinetics experiments (data not shown), I was not able to perform rigorous K_D calculations.

5.2.2.3 Synapsis of in planta and in vitro effector binding analyses of ancHMA mutants

Taking these experiments together, both co-IP and SPR, indicates that the EMVKE region plays an important role in high affinity binding of the AVR-PikD effector by Pikm-HMA. I conclude that the mutations within this region arose as an adaptation to the AVR-PikD effector. The differences between the results obtained by the two methods—in particular differences in binding between ancHMA and AVR-PikD—could be due to a number of factors. As discussed before (**Chapter 4**), these assays differ in sensitivity and experimental design; in the co-IP experiments HMA is presented in the context of a full-length protein, whereas the SPR experiments were conducted using the HMA domain only. It is possible that other domains of the Pikm-1 receptor prevent transient interaction between ancHMA and AVR-PikD observed in vitro or that such

interactions simply cannot be captured in co-IP experiments. Differences between in vitro and in planta results have been previously observed for Pikm-1, which despite moderate interaction with the AVR-PikC effector in SPR experiments does not associate with this effector in co-IP (De la Concepcion et al., 2019).

Both the co-IP and SPR results highlight the key role that the ANK>VKE mutations played in the evolutionary transition of the Pikm-HMA towards strong binding of AVR-PikD. These results are consistent with previous studies of Pikm-HMA (De la Concepcion et al., 2018) and other Pik1–integrated HMA domains. For instance, the NK>KE mutations—corresponding to Lys-262 (K) and Glu-263 (E) of the VKE residues—have been previously shown to be sufficient to confer broader effector-recognition spectrum compared to Pikp-1 (De la Concepcion et al., 2019).

Due to discrepancies between the two methods, the role of the most ancient MK>EM mutations in the evolutionary path of Pikm-HMA remains unclear. Both residues are part of the interface 1, which is considered a minor component of the Pikm-HMA–effector interaction, with a side chain of Met-189 contributing to hydrophobic interactions with AVR-PikD$^{Ile\text{-}49}$ (De la Concepcion et al., 2018). It is therefore possible that these residues have additive effect on AVR-PikD binding. Alternatively, MK>EM mutations could have acted as permissive mutations, opening up the Pikm-HMA mutational path towards subsequent ANK>VKE mutation. However, this seems unlikely, given that aforementioned Pikp-1 with expanded effector recognition was engineered in the MK background (De la Concepcion et al., 2019). Future studies using ancHMA$_{MKVKE}$ mutant may provide a definitive answer on the significance of these early substitutions—if they play a role in binding potentiation or perhaps epistasis.

Interestingly, the EMVKE substitutions are present in only a handful of Pik-1 alleles and orthologues (**Chapter 3**), namely Pik*-1 (Zhai et al., 2011), Pikm-1 (Ashikawa et al., 2008), and Piks-1 (Jia et al., 2009), all of which were cloned from different domesticated rice cultivars, Kusabue, Tsuyuake, and Katy, respectively. The Pik*-1 and Pikm-1 alleles have been proposed to be young alleles that emerged following rice domestication (Kanzaki et al., 2012; Zhai et al., 2011). Indeed, Pikm-1 and Pik*-1 exhibit only eight amino acid polymorphisms—located primarily in the HMA region—but no synonymous changes, further demonstrating that they diverged very recently. This raises the question

whether the EMVKE polymorphisms arose as an adaptation to the emergence of the rice blast lineage and its AVR-PikD effector.

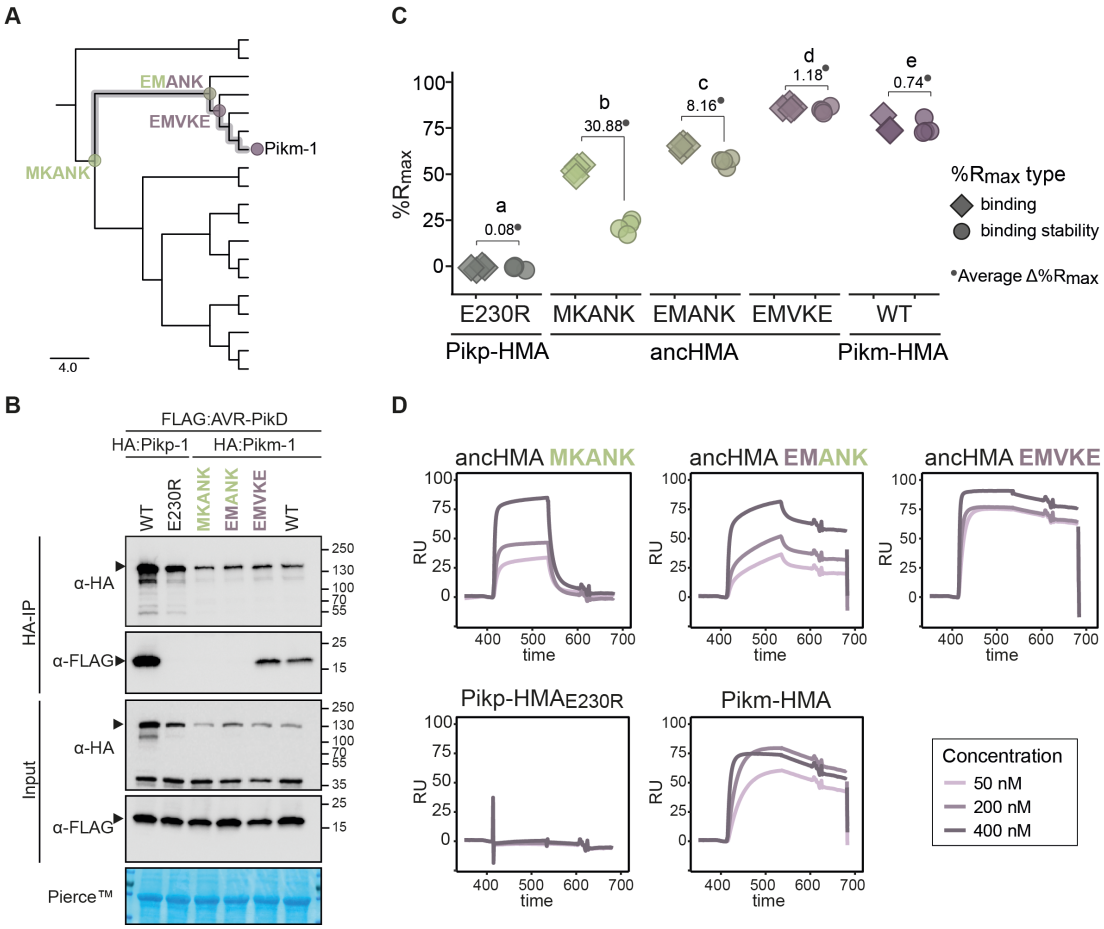


Figure 5.2 ANK>VKE substitutions were essential for Pikm-HMA adaptation towards high-affinity binding to AVR-PikD

(**A**) Schematic representation of the neighbour joining tree of the HMA domains from different *Oryza* species. The scale bar indicates the evolutionary distance based on number of base substitutions per site. Historical substitutions in the EMVKE region acquired over the course of Pikm-HMA evolution are shown next to the corresponding nodes. The mutations are colour-coded to match the ancestral (green) and present-day (purple) states. (**B**) Co-immunoprecipitation experiment illustrating in planta association of FLAG-tagged AVR-PikD with Pikm-1 and Pikm-1:ancHMA constructs featuring HA tag (labelled above). Wild type (WT) HA:Pikp-1/HA:Pikm-1 and HA:Pikp-1$_{E230R}$ constructs were used as positive and negative controls, respectively. Proteins obtained by co-immunoprecipitation using HA resin (HA-IP) and total protein extracts (Input) were immunoblotted with the appropriate antisera depicted on the left. The arrowheads on the left indicate expected band sizes. Rubisco loading control was performed using Pierce™ staining. The experiment was independently repeated at least three times; the replicates of the experiment are shown in **Figure A.4.2**. (**C**) Plot »

Figure 5.2 ANK>VKE substitutions were essential for Pikm-HMA adaptation towards high-affinity binding to AVR-PikD (continued)

illustrating calculated percentage of the theoretical maximum response (%R_{max}) values for interaction of HMA analytes, labelled below, with HIS:AVR-PikD ligand, determined using Surface Plasmon Resonance (SPR). %R_{max} was calculated assuming one-to-one binding model for Pikm-HMA and ancHMAs, and two-to-one (HMA-to-effector) for Pikp-1_{E230R}. The values were normalized for the amount of ligand immobilized on the NTA-chip. The chart summarises results obtained for HMA analytes at 200 nM concentration from five independent experiments, with all the data points represented as diamonds ('binding') or circles ('binding stability'). Three different concentrations of analytes (400 nM, 200 nM, 50 nM) were tested; results for 400 nM and 50 nM concentrations are shown in **Figure A.4.5**. Average Δ%R_{max} (•) values indicate absolute differences between average values for ‚binding' and 'biding stability' for each sample. Statistical differences among the samples were analysed with Tukey's honest significant difference (HSD) test ($p < 0.01$). P-values for all pairwise comparisons are presented in **Table A.4.2**. (**D**) The SPR sonsograms for association and dissociation of the AVR-PikD and HMA proteins in multicycle kinetics experiment, corresponding to the data used in the panel C. His-tagged AVR-PikD was immobilised on the sample cell, giving a response level of 99 ± 33 response units (RU). The replicates of the experiment are presented in **Figure A.4.6**.

5.2.3 Structural determinants of the gain-of-binding phenotype

To better understand the structural determinants of the evolution of high effector binding affinity in Pikm-1, I set out to resolve a crystal structure of $ancHMA_{EMVKE}$ in complex with AVR-PikD. Two-step protein purification from *E. coli* was performed by Adeline Harant from The Sainsbury Laboratory (Norwich, UK). Using purified protein complex, I set up a wide range of commercially available crystallisation screens and optimisations of promising conditions. Crystals, grown in 20% PEG 3350 and 0.1 M potassium phosphate, were tested for X-ray diffraction at Diamond Light Source (Didcot, UK). Although the crystals diffracted to ~1.8 Å resolution (data not shown), poor data quality precluded structure refinement. As an alternative approach, I analysed the significance of the EMVKE region among previously published structures: Pikm-HMA–AVR-PikD structure featuring EMVKE polymorphisms and Pikp-HMA–AVR-PikD carrying LKANK residues reminiscent of the MKANK amino acids present in the ancHMA (De la Concepcion et al., 2018). In both cases lysin 262 is a major binding determinant forming hydrogen bonds or salt bridges with AVR-$PikD^{Glu-53}$ and AVR-$PikD^{Ser-72}$ docking in the HMA-binding pocket (**Figure 5.3**). The position of the lysin differ by one amino acid causing slight differences in the conformation of the

adjacent residues and subsequent alteration of the orientation of the HMA backbone between those two Pik-HMA domains. As described by De la Concepcion et al. (2018), this slight change results in a tighter interaction between Pikm-HMA compared to Pikp-HMA with AVR-PikD. In addition, the EMVKE residues cooperatively facilitate changes in the electrostatic potential generating an exposed hydrophobic surface in Pikm-HMA (absent in Pikp-HMA), which might further contribute to interaction with AVR-PikD, for instance by complex stabilisation. Notably, the homology model of the ancHMA in complex with the AVR-PikD effector—built based on structure of Pikm-HMA (**Chapter 4**)—failed to predict these minor changes, with the modelled HMA backbone positioned as in Pikm-HMA and Lys-262 extending away from the effector. Given that the homology modelling is known to suffer from template-bias and often fails to correctly predict unstructured regions, I concluded that the modelled structure for this particular region is most likely incorrect. Indeed, Pikh-HMA and Pikp-HMA carrying Asn-261-Lys and Lys-262-Glu mutations have been shown to adopt Pikm-like conformation at the interaction interface 3 (De la Concepcion et al., 2020, 2019), confirming that the lysin residue is a major determinant of the structural confirmation within this region. I propose that MKANK residues of the ancHMA form broadly similar interactions with AVR-PikD as those observed for the LKANK residues present in Pikp-HMA. Pikp/Pikm comparison can therefore serve as a reliable proxy for dissecting the overall relevance of the evolution of the EMVKE region of the HMA domain.

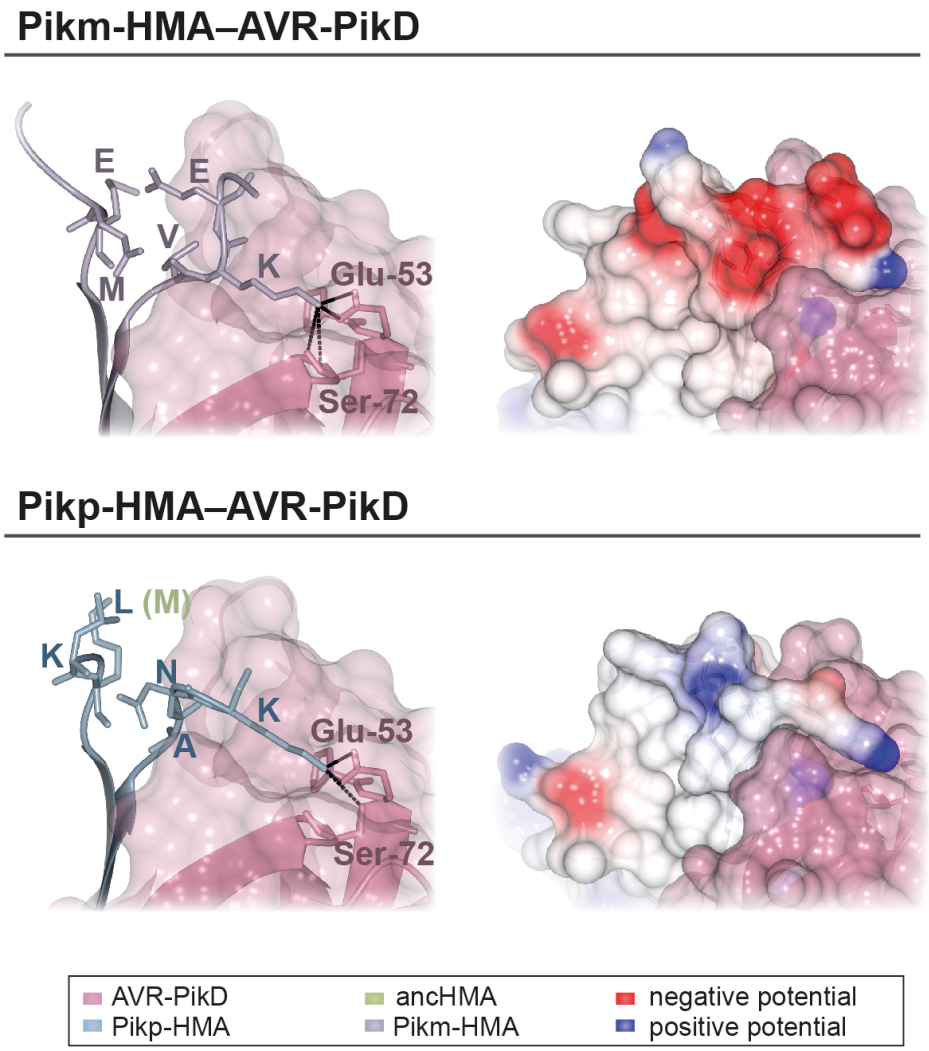


Figure 5.3 Structural differences between Pikm-HMA and Pikp-HMA within the EMVKE region.

Schematic representation of Pikm-HMA (purple) and Pikp-HMA (blue) in complexes with the AVR-PikD effector (pink). The molecules in the left panels are presented as ribbons with selected side chains shown as sticks and labelled; the colours of the residue labels match colours of the respective molecules. The molecular surface of AVR-PikD (pink) is also shown. Dashed lines stand for hydrogen or disulfate bonds formed between the two molecules. The EMVKE residues are labelled with single-letter amino acid symbols—Gln-188 (E), Met-189 (M), Val-260 (V), Lys-261 (K), and Gln-262 (E) for Pikm-HMA; and Leu-187 (L), Lys-188 (K), Ala-260 (A), Asn-261 (N), and Lys-262 (K) for Pikp-HMA; the ancHMA domain carry methionine (M, green) in the structurally equivalent position to Leu-187 of Pikp-HMA. The right panels present protein molecular surfaces, with the surfaces of the HMA molecules coloured according to their electrostatic potential. Structures adapted from: De la Concepcion et al. 2018.

5.2.4 Residues located in the β4–α1 and α2 –β4 loops are responsible for the Pikm:ancHMA autoactivity

To further understand the evolution of the Pikm-1–integrated HMA domain and its potential evolutionary constraints, I performed hypertensive response (HR) experiments. I took advantage of the Pikm-1:ancHMA chimeras (ancHMA$_{EMVKE}$, ancHMA$_{FFE}$, ancHMA$_{IVDPM}$, ancHMA$_{STSN}$, ancHMA$_{VH}$) generated for the purposes of the co-IP experiments, and tested their activity by transiently co-expressing each of the constructs with Pikm-2 in *Nicotiana benthamiana*. As observed with Pikp-1:ancHMA fusions (**Chapter 4**), Pikm-1:ancHMA exhibited strong autoactivity phenotype when co-expressed with Pikm-2 (**Figure 5.4**; **Figure A.4.7**). This autoactivity was compromised in two of the ancHMA mutants, Pikm-1:ancHMA$_{IVDPM}$ and Pikm-1:ancHMA $_{FFE}$. Western blot analyses indicated that this loss-of-autoactivity was not due to reduced protein level (**Figure A.4.1**).

I further co-expressed each of the Pikm-1:ancHMA mutants with both Pikm-2 and AVR-PikD and discovered that while Pikm-1:ancHMA$_{FFE}$ was not responsive to AVR-PikD, co-expression of Pikm-1:ancHMA$_{IVDPM}$ with the effector resulted in localised cell death. The HR triggered by remaining constructs was not affected by the AVR-PikD effector.

The IVDPM amino acids reside in the parallel loops separating β1 from α1 (β1–α1) and α2 from β4 (α2–β4). Remarkably, this region is a structural equivalent of the AMEGNND region of the Pikp-1 HMA domain, which I showed to compromise the autoactivity of Pikp-1:ancHMA (**Chapter 4**). This observation suggests that the autoactivity phenotype for both the Pikm-1:ancHMA and Pikp-1:ancHMA fusions are likely induced by a similar mechanism and could involve intra- or intermolecular interactions with other Pik domains or interactions with metal ions.

Strikingly, Pikm-1:ancHMA$_{IVDPM}$ triggered HR upon co-expression with AVR-PikD, despite the fact that I was unable to detect any significant association with these constructs in the co-IP assays (**Figure A.4.1**). Considering that the co-IP experiments were performed in the absence of Pikm-2, it is possible that the helper plays a role in facilitating the AVR-PikD–Pikm-1 interaction, such as it enables formation of a stable and robust complex. Notably, despite robust HR, overall strength of association between AVR-PikD and the full-length Pikm-1 constructs, including wild type Pikm-1, was quite low.

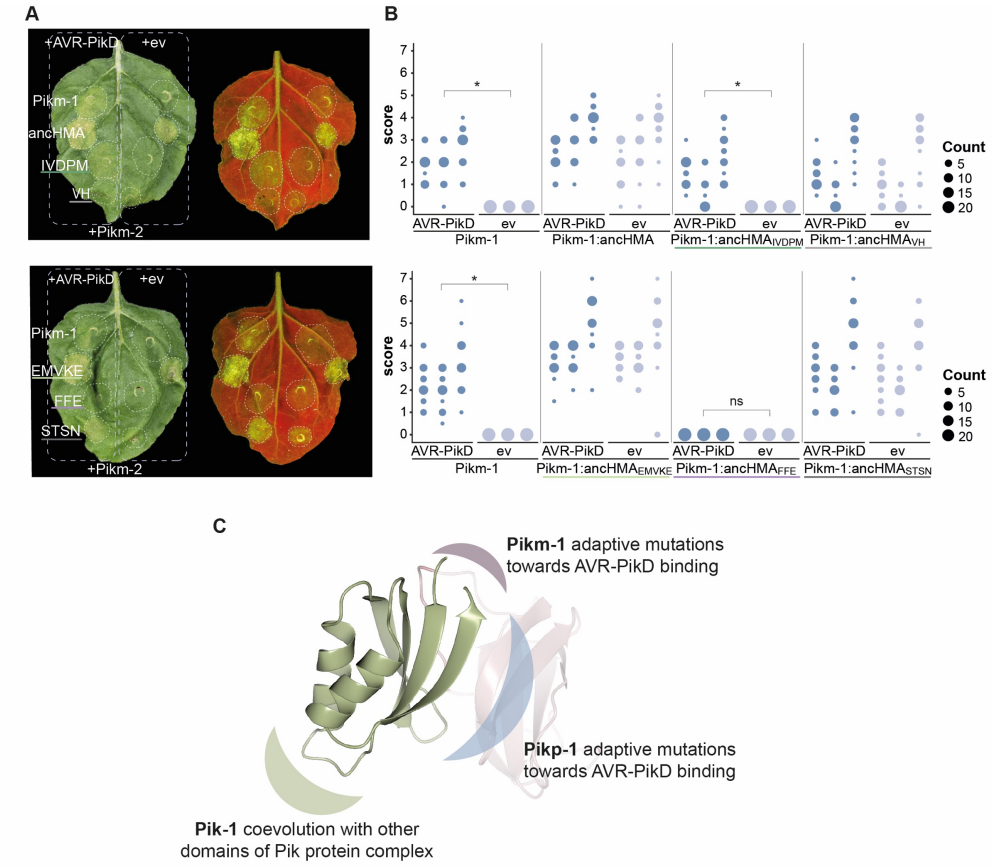


Figure 5.4 Substitutions within the FFE and IVDPM regions compromise autoactivity of Pikm-1:ancHMA fusions

Cell death assay after transient co-expression of Pikm-1:ancHMA chimeras with AVR-PikD/empty vector (ev). Empty vector (ev) was used as a negative control. (**A**) Representative *Nicotiana benthamiana* leaves infiltrated with appropriate constructs (labelled next to the infiltration spot) were photographed five days after infiltration under day light (left) and UV light (right). (**B**) Hypersensitive Response (HR) was scored at five days post-infiltration. The results are presented as a dot plot where a size of a dot is proportional to the number of samples (count) with the same score within the same biological replicate. The experiment was independently repeated at least three times with 21–24 internal replicates; the columns within tested conditions (labelled on the bottom) correspond to results from different biological replicates. The differences in HR were analysed and visualised using the estimation methods (MacLean, 2019) as presented in **Figure A.4.7**. Statistical significance for selected samples is marked with asterisk (*); ns: not significant. (**C**) A model of evolution of the Pik-1–integrated HMA domain (green) with regions comprising of adaptive mutations marked.

To add another level of complexity, Pikm-1:ancHMA$_{FFE}$ also showed loss-of-autoactivity in the HR assays yet no response to AVR-PikD. The FFE amino acids are located in close proximity to the IVDPM residues that stretch towards the

effector-binding interface. As a consequence, these residues could potentially modulate effector binding, explaining marked differences between Pikm-1:ancHMA$_{FFE}$ and Pikm-1:ancHMA$_{IVDPM}$ in the HR experiments. Another explanation could come from the position of the side chains of the mutated residues. The FFE residues face towards the core structure of the HMA domain, and could therefore impair the overall structure of the HMA domain, in contrast to the IVDPM mutations where side chains are mainly surface-exposed.

To sum up, I discovered two adjacent regions within the ancHMA that abolish Pikm-1:ancHMA autoactivity. Future research should determine the mechanism underlying Pik activation and the role that the regions identified in this study play in the autoactivity. Finally, the knowledge gained from these experiments could be leveraged to address questions about the correlation between the strength of immune response and effector binding, particularly those mediated by the EMVKE region.

5.3 Conclusions

In this chapter, I reconstructed the evolutionary trajectory of the Pikm-1–integrated HMA domain and functionally characterised residues responsible for the adaptation towards high-affinity binding and recognition of the rice blast effector, AVR-PikD. These results fit well within the context of previous research performed with Pikm-HMA, and provide a better understanding about the role of selection in adaptive evolution of Pikm-1. Interestingly, the adaptive mutations to AVR-PikD identified in this study were located at a different interface compared to Pikp-HMA (**Chapter 4**), indicating that although Pikp-1 and Pikm-1 receptors evolved to bind AVR-PikD, they followed different evolutionary pathways to do so. In addition, functional analyses of the ancHMA mutants helped to determine the approximate region responsible for the autoactivity of the Pikm-1:ancHMA fusion, which provide a useful framework for future research to address questions about Pik activation.

Chapter 6: Discussion

Despite remarkable advances in the field of NLR biology, there remains a significant gap in our understanding of how these receptors have evolved and adapted to ever-changing pathogens. NLRs that have acquired novel recognition specificities through the fusion of extraneous domains can serve as an excellent experimental system to address some of the prevailing questions. In this thesis, I used the rice Pik-1/Pik-2 system to test hypotheses about adaptive evolution of NLRs and their integrated domains. In chapter 3, I described the rich genetic diversity of the *Pik-1*/*Pik-2* genes in grasses and discovered that they likely derived from a single ancestral gene pair that emerged before the radiation of the BOP and PACMAD clades. Further phylogenetic analyses of Pik orthologues revealed that the HMA domain integrated into Pik-1 before the emergence of Oryzinae and has been subject to the strong diversifying selection since. In chapters 4 and 5, I took advantage of the detailed knowledge of the structure and function of the Pik-1–integrated HMA domain and explored how the AVR-PikD effector from *M. oryzae* has shaped the evolutionary trajectories of this domain in Pikp-1 and Pikm-1 alleles of Pik-1. I performed ancestral sequence reconstruction and biochemically characterised the resurrected HMA, revealing that these different allelic variants of Pik-1 convergently evolved to recognise AVR-PikD. Using in vitro and in planta assays, I functionally characterised the two regions in Pik-1 HMAs that independently evolved high affinity AVR-PikD binding from a weaker ancestral state. I further reconstructed stepwise mutations that led to these adaptations, shedding light on the mechanisms and dynamics of protein evolution.

6.1 Convergent evolution of the Pik-1 alleles—different biochemical solutions to the same problem

Convergent evolution is broadly defined as the independent evolution of similar phenotypic features in distinct lineages (Blount et al., 2018; Losos, 2011). Similar phenotypes can emerge by common ancestry, or through constraints in the space of possible phenotypes that can increase in a population via either drift or selection (Washburn et al., 2016). In general, recurring convergence among biological systems has

long been considered as evidence for adaptation to the same selection pressure (Rokas and Carroll, 2008; Zou and Zhang, 2015). Given their unique evolutionary dynamics and extreme selection, plant–microbe systems are exceptional in the number of examples of rapid and repeated evolution of similar adaptive traits. Patterns of convergent evolution among NLRs can be detected at multiple levels. Many NLRs display marked similarities in genetic architecture, mechanisms of activation, and interconnectivities within plant NLRomes. For instance, there is substantial evidence indicating that dependence on a helper NLR evolved more than once among different plant families, as illustrated by immunoreceptor networks and NLR pairs (Adachi et al., 2019b; Tamborski and Krasileva, 2020; Upson et al., 2018). Moreover, unrelated NLR receptors sometimes display functional convergence—they can share similar specificities and recognise the same pathogen effectors (Ashfield et al., 2004; Kessens et al., 2014; Selote and Kachroo, 2010). Independent evolution of different Pik-1 alleles represents yet another type of convergent evolution in which closely related sequences evolve similar recognition specifies by acquiring distinct mutations. Nonetheless, the molecular bases of functional transitions in NLR adaptive evolution remain poorly understood and comprehensive evolutionary reconstructions based on robust phylogenetic frameworks are lacking.

In this thesis, I address this gap by investigating the biochemical basis of the step-by-step adaptive evolution of Pikp-1 and Pikm-1 towards high affinity binding to the AVR-PikD effector from a weaker ancestral form. I showed that the two alleles of the Pik-1 receptor have evolved to produce similar phenotypic outcomes, despite having undergone different evolutionary trajectories to do so. It is remarkable that they evolved to recognise the AVR-PikD effector by acquiring independent mutations in two distinct regions within the HMA domain; Pikp-1 evolved to interact with AVR-PikD with high affinity in the LVKIE region located in the interface 2 (**Chapter 4**), whereas Pikm-1 acquired adaptive mutations in the EMVKE region of the interface 3 (**Chapter 5**) (De la Concepcion et al., 2018). Overall, these HMA interfaces seem to function in a broadly synergistic yet interchangeable manner, such that weak interaction at one interface can be compensated by strong interaction at a different one. This modularity between different regions of the HMA possibly increases the HMA's capacity for rapid adaptive evolution as it can follow alternative mutational paths to produce similar phenotypic outcomes and counteract rapidly evolving pathogen effectors. This is

significant since multiple studies on protein evolution have demonstrated that in general there are only limited numbers of evolutionary trajectories that a given protein can follow, owing to constraints imposed by protein function, physical properties, and geometry (Blount et al., 2018; Rokas and Carroll, 2008; Weinreich et al., 2006). The structural plasticity of the HMA domain, together with further functional compartmentalisation discussed in **section 6.2,** can therefore act as a key evolutionary innovation that expands the adaptive landscape of the Pik-1 receptor.

6.2 NLR-ID: an evolutionary innovation providing balance between robustness and evolvability?

One of the hallmarks of all biological systems is the ability to persist in the face of changing conditions (Lenski et al., 2006). Evolvability—the ability to generate phenotypic variation that is both heritable and adaptive—is therefore a fundamental feature of biological systems (Fares, 2015). At the same time, phenotypic plasticity should not compromise the organism's performance and persistence (survivability) (Palmer and Feldman, 2012)—biological systems must somehow accommodate genetic changes, which is often defined as mutational robustness. At the molecular level, mutational robustness can be achieved through gene duplication, with the duplicated copy released from the functional constraints, enabling accumulation of novel mutations and potential sub- or neofunctionalization (Hartman et al., 2001; Tokuriki and Tawfik, 2009). This mechanism constitutes a recurrent theme in the evolution of plant immune receptors, with many examples of NLR families that derived from a single gene through duplication and diversification (Kim et al., 2017b; Shao et al., 2016; Stam et al., 2019; Wu et al., 2017). NLRs carrying an integrated domain seem to follow an alternative evolutionary strategy, which does not involve massive diversification. Instead, in the few cases that have been studied, NLR-IDs appear to evolve novel specificities through recycling ancient variation, a model of evolution that has been proposed for a number of proteins, including plant NLRs (Bridgham, 2006; Huang et al., 2009; Karasov et al., 2020; Wicker et al., 2007).

I propose that fusion with extraneous domains and multilayer functional compartmentalisation within NLR-ID proteins provide a robust evolutionary mechanism to increase the adaptive landscape of these NLRs without compromising their activities.

Firstly, integrated domains might act as malleable platforms to adapt to new pathogen effectors by accommodating an accelerated mutational rate without dramatically perturbing the remainder of the NLR protein. This is reminiscent of enzymes where conformationally dynamic active sites are commonly thought to provide functional promiscuity while also maintaining structural integrity (Tokuriki and Tawfik, 2009). The levels of diversifying selection detected in the HMA domain of Pik-1 described in this thesis are astonishing, highlighting the evolutionary potential of the integrated domains (high evolvability) (**Chapter 3**). This is consistent with previous reports describing the integrated HMA domain as the most diverse among Pik-1 domains (Costanzo and Jia, 2010). Moreover, having a domain that's solely responsible for effector recognition may release other domains from the pressure of diversification and reduces the risk of compromising or mis-regulating NLR activity (Césari, 2018). The selection tests performed in this thesis are consistent with the model in which the HMA of Pik-1 has diversified more extensively than the canonical NB-ARC domain, likely reflecting rapid coevolution with plant pathogens versus purifying selection. Similar observations have previously been made in a number of plant NLRs, where individual domains can display patterns of asymmetrical evolution and distinct rates of positive selection, suggesting that NLRs evolve in a modular fashion (Kuang et al., 2004; Maekawa et al., 2019; Prigozhin and Krasileva, 2020; Read et al., 2020; Seeholzer et al., 2010). Finally, coupling with the Pik-2 helper likely provides yet another mechanism of functional compartmentalisation, further enhancing the evolvability of the sensor by freeing it from the constraint of executing the hypersensitive cell death.

NLR-IDs are widely present across species from diverse plant families—including early diverging species such as mosses (Gao et al., 2018)—which indicates that domain integration is a common mechanism of evolution of plant immune receptors, and might indeed play an important role in NLR diversification. Furthermore, as in *Pik-1*/*Pik-2*, NLR-IDs are commonly arranged in a head-to-head orientation with phylogenetically unrelated NLR genes (Ashikawa et al., 2008; Fujisaki et al., 2017; Okuyama et al., 2011; Saucet et al., 2015). The function and activities of many of these NLR pairs remain elusive but, similar to Pik-1/Pik-2 pairs, some display clear patterns of coevolution, suggesting that they might be functionally linked (Van de Weyer et al., 2019). It would be fascinating to see whether the evolutionary patterns described in this work, namely intramolecular

functional compartmentalisation, can be observed in other NLR-IDs and their partners, and to what extent the mechanisms of evolution are shared across this class of receptors.

6.3 Has the integrated HMA evolved to recognise the AVR-PikD effector following the blast fungus jump to domesticated rice?

It is intriguing that the integrated HMA domain of Pik-1 has acquired the capacity to strongly bind AVR-PikD relatively recently considering its long evolutionary history. In this thesis, I determined that the ability to bind this effector was acquired rapidly and, at least in the case of Pikp-HMA, was not contingent on historical mutations within this domain (**Chapters 4** and **5**). The high sequence divergence and elevated mutation rates precluded rigorous dating of the emergence of the Pikp- and Pikm-HMA adaptations towards AVR-PikD binding. However, the low level of total nucleotide polymorphisms among closely related Pik alleles—in particular, the extremely low number of synonymous substitutions among Pikp- and Pikm-related alleles—points to a recent emergence of the adaptative mutations. I put forward the hypothesis that these new blast resistance specificities arose in an agricultural context after the host jump of the blast fungus to domesticated rice.

Crop domestication and spatial proximity of closely related plant species commonly facilitate pathogen emergence and host jump events (McCann, 2020). The high frequency of dense populations of largely genetically uniform hosts represent a fertile ground for pathogen emergence, specialization, and genetic divergence. Although we do not have a complete picture of the many mechanisms that underpin host specialisation, there is substantial evidence indicating that variation in effector repertoires can act as specificity determinants and contribute to host jumps as well as adaptation and specialization on the new host (Dong et al., 2014; Inoue et al., 2017; Raffaele et al., 2010; Yoshida et al., 2016).

Rice blast is thought to have originated through a host jump of *Magnaporthe oryzae* from foxtail millet to rice around 7,000–9,000 years ago (Couch et al., 2005; Latorre et al., 2020). This jump and subsequent host adaptations have been attributed to extensive gains and losses of effector genes associated with transposable elements (Yoshida et al., 2016). *AVR-Pik* exhibits lineage-specific presence/absence polymorphisms and is present exclusively in rice-infecting isolates, which suggests that it is a hallmark of the

rice-infecting lineage. Furthermore, given its widespread distribution among rice blast isolates (Langner et al., 2020; Latorre et al., 2020), AVR-Pik presumably plays an important role in rice infection—and perhaps even host specialisation—albeit its exact role during infection remains confounded by the occurrence of other HMA targeting effectors.

It is interesting to speculate that the emergence of the AVR-Pik effector and its widespread occurrence in rice-infecting populations of the blast fungus created the ecological context that led to Pik neofunctionalization towards recognition of this new pathogen threat. Different rice populations must have independently encountered pathogens carrying this effector, leading to the independent emergence of the Pikp and Pikm subclades. This can be further interpreted as evidence of the prevalence of AVR-Pik in the pathogen populations and its potential as an HMA-binding host-translocated effector to become an AVR target for the HMA-containing Pik receptor. It is noteworthy that while classic gene neofunctionalization usually involves gene duplication followed by gain-of-function mutations (Boucher et al., 2014; Voordeckers et al., 2012), the Pik alleles evolved new specificity through direct sequence diversification, implying that they had not been constrained by selection for a different ligand. Whether ancestral Pik was involved in recognition of a dispensable effector, or acted as a temporarily 'out-of-use' reservoir of genetic diversity, can only be postulated (see below).

6.4 What was the function of the ancestral HMA domain?

In this work, I uncovered the evolutionary history of the Pik-1–integrated HMA domain and hypothesised that the HMAs of both the Pikp-1 and Pikm-1 receptors evolved to bind AVR-PikD with high affinity as a result of selection pressure imposed by the newly emerged rice blast lineage of *M. oryzae*. Yet Pik-1 and its integrated HMA domain appear to be much more ancient than rice blast, inviting the question about the role of the ancestral state of the HMA domain. More specifically: prior to the emergence of AVR-PikD binding, which selection pressure drove HMA evolution?

One possibility is that the ancestral million-years-old HMA recognized a different pathogen effector. Considering that HMA proteins are known to interact with structurally similar ligands, notably MAX-effectors (de Guillen et al., 2015; De la Concepcion et al.,

2020; Guo et al., 2018; Ose et al., 2015), one may speculate that the ancHMA recognised one such effector and was therefore structurally preadapted to bind AVR-PikD which evolved millions of years later. MAX effectors form an ancient effector family present across multiple blast lineages and other phylogenetically distant pathogen species (de Guillen et al., 2015; Petit-Houdenot et al., 2020), and a number of MAX-type effectors have known virulence functions and many act as avirulence factors (Fujisaki et al., 2015; Kang et al., 1995; Kanzaki et al., 2012; Okuyama et al., 2011; Singh et al., 2016; Tosa et al., 2005). It is possible that multiple MAX-effectors shaped the evolution of the integrated Pik-HMAs, and that the HMA integration itself was also driven by yet another ancient MAX effector. Testing ancHMA for interaction with this class of proteins could challenge these hypotheses and provide indirect evidence that MAX effectors played a role in the evolution of the Pik-1–integrated domain. My colleague Yohann Petit from the Kamoun Lab has performed a preliminary yeast 2-hybrid screen of the ancHMA against a library of predicted MAX-type effectors from *M. oryzae* and identified putative ancHMA interactors, revealing the possible promiscuity of this domain towards MAX proteins.

I propose that following integration, the HMA domain has served as an interaction module to recognize sequence-unrelated effectors. This model paints a complex picture of the macroevolutionary dynamics of plant–microbe interactions. One could argue that NLRs caught in pairwise arms races (one NLR recognising one effector) are more likely to be short-lived, whereas NLRs entangled in diffuse evolution (functioning against multiple effectors and/or multiple pathogens) could be useful over longer timescales (Karasov et al., 2014). It is remarkable that the *Pik-1* gene carrying the integrated HMA domain has been maintained in rice populations for millions of years. This points to a highly successful evolutionary strategy for generating long-lived resistance, with potential promiscuity of the ancHMA towards different MAX-effectors at the centre of this model.

This model is also consistent with the view that effectors from different pathogens have repeatedly evolved to converge on the same host signalling components and pathways (Macho and Zipfel, 2015). Integration of effector hubs into immune receptors might therefore increase robustness of the immune response against multiple effectors or pathogens, and underly the longevity of those receptors.

In the future, the discovery of pathogen effectors, from *M. oryzae* or otherwise, that bind ancHMA will help challenge this model and test the view that Pik-1 first evolved to target a different effector but then switched or expanded its activity to bind and respond to AVR-PikD following the emergence of rice blast lineage of *M. oryzae*.

6.5 Autoactivity: a major constraint in the evolution of NLR-IDs

While investigating the evolutionary dynamics between the Pikp-1–integrated domain and the AVR-PikD effector, I discovered that many of Pik-1:ancHMA fusions were autoactive, adding another layer of complexity to HMA evolution. I found that Pikp-1:ancHMA as well as Pikm-1:ancHMA triggered spontaneous cell death when co-expressed with Pikp-2 (**Chapter 4**) or Pikm-2 (**Chapter 5**), respectively. I further mapped the region responsible for the autoactivity to two parallel loops, separating β1 from α1 and α2 from β4, and the surrounding residues. Structurally equivalent regions underpin the autoactivity phenotypes for both Pikm-1:ancHMA and Pikp-1:ancHMA, indicating that autoactivity is likely induced by a similar mechanism that could involve intramolecular interactions with other Pik-1 domains, intermolecular interactions with Pik-2, or interactions with metal ions.

Intramolecular incompatibility has previously been reported for several NLR chimeras and has been linked to an autoimmunity phenotype (Lukasik-Shreepaathy et al., 2012; Qi et al., 2012; Rairdan and Moffett, 2006; Slootweg et al., 2013; Wang et al., 2015). Relatedly, research on the RRS1-integrated WRKY domain has suggested that the integrated domain contributes to maintaining the immunoreceptor complex in an inactive state (Ma et al., 2018). Given the marked differences between the integrated domains of Pik-1 and RRS1, and their distinct integration positions, it is unlikely that the same suppression mechanism applies to these proteins. In addition, Pik-1/Pik-2 appear to function through an intermolecular cooperation model that is distinct from the activation model proposed for RRS1/RPS4 (Zdrzałek et al., 2020). Although the exact mechanism of Pik-mediated immunity remains obscure, it is possible that different domains of the Pik-1 alleles are involved in intramolecular interactions that keep the complex in the correct conformation, switching between active ('ON') and inactive ('OFF') states, depending on the stimuli. Mismatching domains from different evolutionary timepoints

could disrupt these proposed interactions, indicating that fine-tuned intramolecular cooperation stems from domain coevolution. It is conceivable that, in keeping with this model, over the course of Pik-1 evolutionary history, the diversification of the HMA sequence must have been balanced by compensatory mutations in other parts of the protein that prevented spontaneous activation.

It is also possible that ancestral HMA might be involved in mismatched interactions with Pik-2 or other, yet undiscovered, signalling components, and cause autoimmunity reminiscent of hybrid necrosis (Alcázar et al., 2009; Barragan et al., 2019; Bomblies et al., 2007; Bomblies and Weigel, 2007; Chae et al., 2016; Deng et al., 2019; Li et al., 2020; Tran et al., 2017; Yamamoto et al., 2010). Previous research has shown that co-expression of Pikp-1 with Pikm-2 results in spontaneous cell death in *N. benthamiana*, and this Pik-1/Pik-2 incompatibility has been mapped to the HMA domain (De la Concepcion, 2020), consistent with the hypothesis that HMA can influence sensor/helper interactions. The β1–α1 and α2–β4 loops of the HMA domain described in this work form a flexible and exposed surface, making it possible that they mediate physical interactions with other immune components or indirectly modulate the overall structure of the Pik complex. Notably, many residues located in these regions display extreme or moderate signatures of positive selection (**Chapter 3**), inviting the question about the possible trajectories towards functional combinations.

Incidentally, the residues within the β1–α1 loop are part of a heavy metal-binding site, characteristic of the HMA protein family, which is defined by the MxCxxC motif (De Abreu-Neto et al., 2013; DeSilva et al., 2002). The motif is degenerated in the panel of present-day Pik HMAs, such as Pikp- or Pikm-HMAs, while the ancHMA sequence contains potentially active MxSxxC motif (Banci et al., 2008). This invites an alternative model, in which affinity towards metal molecules could interfere with Pik function, resulting in spontaneous cell death. In line with this model, insertion of non-integrated HMAs into Pikp-1 leads to autoactivity that can be compromised by mutating residues within the metal-binding loop (Maidment et al., 2020). To date, however, there's no evidence that non-integrated rice HMAs or the ancHMA can interact with heavy metals.

Future research should determine the mechanism underlying Pik activation and the role that the regions identified in this study play in the autoactivity. Regardless of the precise mechanism, autoactivity is a liability for the plant and NLR domain integration

can incur considerable fitness costs. This is illustrated by the aforementioned experiments with Pik-1:ancHMA fusions, but also by occasional large deletions within the HMA, as in *Lp*Pik-1 and possibly in N-type *Pik-1* genes, which may have emerged to eliminate autoimmunity (**Chapter 3**). I propose that the risk of autoactivity acts as a strong evolutionary constraint narrowing HMA/NLR mutational pathways. Those constrains should be considered when studying NLR diversification. The knowledge gained from this and future studies could be further utilised in NLR engineering.

6.6 Navigating plant genetic diversity through an evolutionary perspective—reconstruction of the deep evolutionary history of the *Pik* genes

Increasing availability of whole plant genome and transcriptome sequences has revealed an incredible genetic diversity across and within species enabling reconstruction of trait evolution in plant systems (Delaux et al., 2019). Comparative approach performed within an evolutionary framework (phylogenomics) was initiated by developmental biologists, leading to emergence of the evolutionary developmental biology (evo-devo) (Carroll, 2005). Since then, evolution of host genes, gene networks, and entire pathways have become a subject of research in plant symbiotic interactions and more recently in plant–pathogen systems—collectively referred to as evolutionary molecular plant–microbe interactions (evoMPMI) (Upson et al., 2018). Phylogenomics has helped to uncover coevolving pathway components, which have since been functionally validated (Baggs et al., 2020; Delaux et al., 2015, 2014; Radhakrishnan et al., 2020). In the field of NLR biology, a number of genome-mining studies coupled with phylogenetic and functional analyses revealed distinct types of NLRs, many of which have been shown to display unique activities or engage in sophisticated NLR networks (Bailey et al., 2018; Gao et al., 2018; Kroj et al., 2016; Lee et al., 2019; Sarris et al., 2016; Shao et al., 2016; Stam et al., 2019; Van de Weyer et al., 2019; Wu et al., 2017). Other studies have used comparative sequence analyses to investigate conservation patterns among NLR proteins, or NLR homologues, that could be predictive of their critical activities, and further help in formulating hypotheses that could be experimentally tested (Adachi et al., 2020, 2019a; Prigozhin and Krasileva, 2020). Comparative approach has been used to uncover distinct

evolutionary features across NLR domains, including NLR integrated domains (Brabham et al., 2017; Kuang et al., 2004; Maekawa et al., 2019; Prigozhin and Krasileva, 2020; Read et al., 2020; Seeholzer et al., 2010). Finally, an evolutionary perspective could help to identify key events in NLR evolution, such as transition from singletons to NLR pairs or networks (Adachi et al., 2019b), or acquisition of integrated domains (Bailey et al., 2018).

In this work, I reconstructed the evolutionary history of Pik-1 and Pik-2 receptors (**Chapter 3**). I took advantage of publicly available genomes and by means of phylogenetic analyses and recurrent BLAST searches I uncovered rich genetic diversity of *Pik* genes across distantly related species. This helped me to date the emergence of the *Pik* pair to before the split of two major grass lineages: the BOP and PACMAD clades, which corresponds to ~100 to 50 MYA, depending on dating analysis (Hodkinson, 2018). Furthermore, by comparing the phylogeny of plant species with the presence/absence of the HMA domain I was able to estimate that Pik-1 acquired the HMA domain prior the emergence of Oryzinae.

While examining the genetic diversity of *Pik* across grasses, I discovered that the vast majority of *Pik-1* and *Pik-2* orthologues exist as pairs; they are genetically linked in head-to-head orientation and display patterns of coevolution across Poaceae. This suggests that the pairing of *Pik-1* and *Pik-2* occurred prior to the integration of HMA into Pik-1. Tight genetic linkage of paired NLRs, such as Pik-1/Pik-2 (Ashikawa et al., 2008), RGA5/RGA4 (Césari et al., 2013; Okuyama et al., 2011), RRS1/RPS4 (Saucet et al., 2015), or RPP2A/RPP2B (Sinapidou et al., 2004), is thought to facilitate coregulation and coevolution, thereby ensuring proper cooperation between these NLRs and reducing the genetic load caused by autoimmunity (Baggs et al., 2017; Griebel et al., 2014; Wu et al., 2017). Notably, a number of Pik-1 and Pik-2 paralogues are also encoded by adjacent genes—a phenomenon previously observed in wild and cultivated rice (Mizuno et al., 2020). Although the function of those homologues remains largely obscure, it's interesting to speculate that their genetic linkage and coevolutionary history might be a hallmark of cooperation between Pik-1 and Pik-2 receptors, rather than a simple consequence of genetic sweep. Perhaps one of the most compelling questions emerging from these analyses is whether prior the integration Pik-1/Pik-2 had already functioned through cooperation and whether genetic and proposed functional linkage with Pik-2 predisposed Pik-1 for the HMA integration.

Furthermore, I uncovered that while in the majority of species Pik-1 and Pik-2 are encoded by single-copy genes, members of the Pooideae and Panicoideae subfamilies frequently encode multiple Pik-1 or Pik-2 receptors, with wheat carrying the highest number of orthologues. Many of those paralogues are genetically linked in head-to-head orientation, but some exist as genetic singletons. Such expansions could broaden NLR genetic diversity within those species by promoting gene diversification and subsequent sub- or neofunctionalization, commonly associated with NLR proteins (Kim et al., 2017b; Wu et al., 2017). It is possible that in these species linked and unlinked *Pik-1* and *Pik-2* genes operate as a receptor network as described for other NLRs (Castel et al., 2019; Wu et al., 2017).

Comparative sequence analyses coupled with calculation of substitution rates revealed other unique features across *Pik* genes. As mentioned above, estimating rates of synonymous substitutions across *Pik-1* and *Pik-2* genes showed that genetically linked *Pik* genes have the same molecular age, providing evidence that *Pik-1* and *Pik-2* have been coevolving for the last 50 to 100 million years. Further tests revealed distinct molecular features across the NB-ARC and HMA domains of Pik-1. Unlike the NB-ARC domain, the Pik-1–integrated HMA domain stood out for displaying elevated d_N/d_S ratio indicative of positive selection. I further mapped specific regions within the HMA that are likely under positive selection. My hypothesis is that this strong HMA diversification is driven by antagonistic molecular interactions with pathogen effectors.

To summarise, rich genetic diversity of the Pik NLRs and their relative conservation across diverse grass species offered a unique opportunity for comparative analyses and enabled reconstruction of deep evolutionary history of these receptors. An evolutionary perspective provided a framework for drawing the links between NLR sequence, gene structure, and domain architecture, and allowed generating experimentally testable hypotheses.

6.7 Evolutionary approach to studying mechanisms of molecular plant–microbe interactions

Plant–microbe systems are incredible for their evolutionary dynamics that can be studied across multiple timescales (Upson et al., 2018). The molecular arms race between

plant NLRs and pathogen effectors is illustrative of microevolutionary plant–pathogen dynamics with rapid and specific sequence adaptations on both sites. With longer time scales in mind, microbes have shaped the evolution of plants since even before the emergence of land plants—and vice versa. This resulted in refined interactions with increased robustness through redundancy and functional compartmentalization in plant immune systems.

Ever since the cloning of the first NLR gene, there have been tremendous advancements in our understanding of plant immunity and NLR function. The incredible range of molecular mechanisms and models continue to emerge, shedding light on many different aspects of function of plant immune systems. Unfortunately, much of the research in the field of molecular plant–microbe interactions has been performed without the appreciation of how those systems came to be the way that they are. I would argue that much of mechanistic research in this field would greatly benefit from interdisciplinary approaches conducted within a robust phylogenetical and ecological framework.

In this work, I aimed to dissect the physical mechanisms and evolutionary processes by which the rice Pik-1/Pik-2 receptor pair diversified and adapted to sense pathogen effectors. Following extensive phylogenetic analyses, I performed ancestral sequence reconstruction of the Pik-1–integrated HMA domain and conducted a wide range of biochemical assays using resurrected ancHMA. Integration of evolutionary and biochemical approaches enabled me to experimentally challenge the hypothetical models about adaptive evolution of Pik proteins and gain a more comprehensive understanding of selection forces, historical contingency, and evolutionary constraints shaping their properties. The Pik-1/Pik-2 receptor pair transpired as an excellent and amenable model system for investigating several open questions in molecular evolution of NLR proteins, where evolution across genotype–phenotype space was experimentally traceable. I argue that the clear evolutionary framework applied to the research described in this thesis not only helped to provide novel insights, but also added exciting dimension to our mechanistic understanding of the system. Furthermore, ancestral sequence reconstruction—a method that has rarely been used in the field of plant–microbe interactions (Dong et al., 2014; Tanaka et al., 2014; Thornton, 2004; Zess et al., 2019)—has transcended phylogenetic inference to yield more accurate evolutionary models.

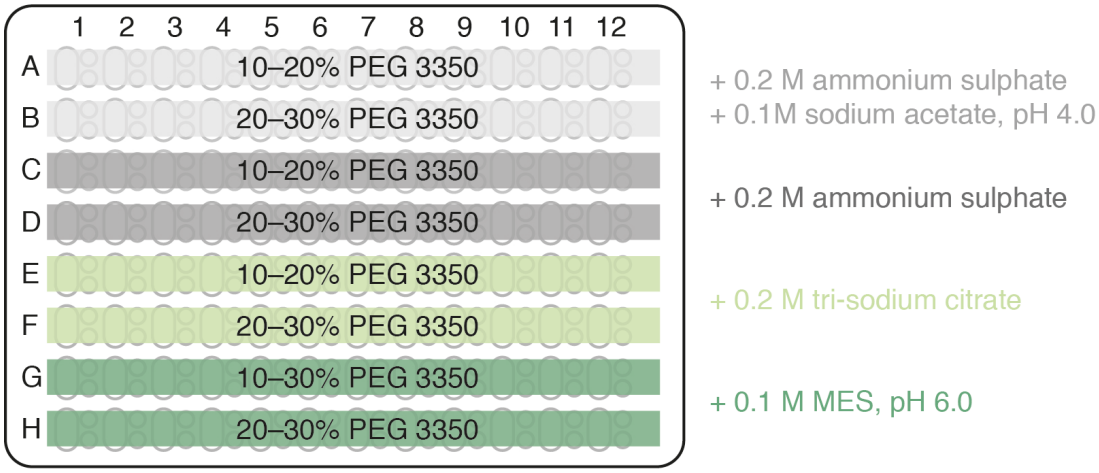
6.8 Concluding remarks and future directions

The research performed as part of this thesis illustrates how mechanistic research structured by a robust evolutionary framework can enhance our understanding of plant–microbe systems. The experiments performed in this work paint a complex picture of the dynamic evolution of NLR-IDs with countless open questions still remaining unanswered. For example, what selected for the integration of the HMA domain into the Pik-1 receptor, and what type(s) of effector(s) shaped HMA evolution prior the emergence of the AVR-PikD effector? What are the determinants of Pik-1:ancHMA autoactivity; does the spontaneous activation stem from inter- or intramolecular interactions of Pik complex, the ancHMA metal-binding activity, or perhaps an entirely different mechanism? Do Pik-1–integrated HMAs from wild rice species also mediate recognition of pathogen effectors? If so, what are these effectors? What is the role of Pik-1 orthologues that don't carry the integration? Are these active NLRs that mediate resistance against plant pathogens, or do they act as a reservoir of genetic diversity for evolving future resistance specificities? Do genetically linked Pik-1 and Pik-2 orthologues function as pairs across different species? From a broader perspective, do similar evolutionary patterns persist across a wide range of NLR-IDs? To what extent do evolutionary models drawn from studying Pik-1 receptors can be applied to other NLR-IDs? Answering these fundamental questions will help advance our understanding of plant immunity with the appreciation of how those systems evolved, bringing a greater degree of rigour to our understanding of plant–microbe systems.

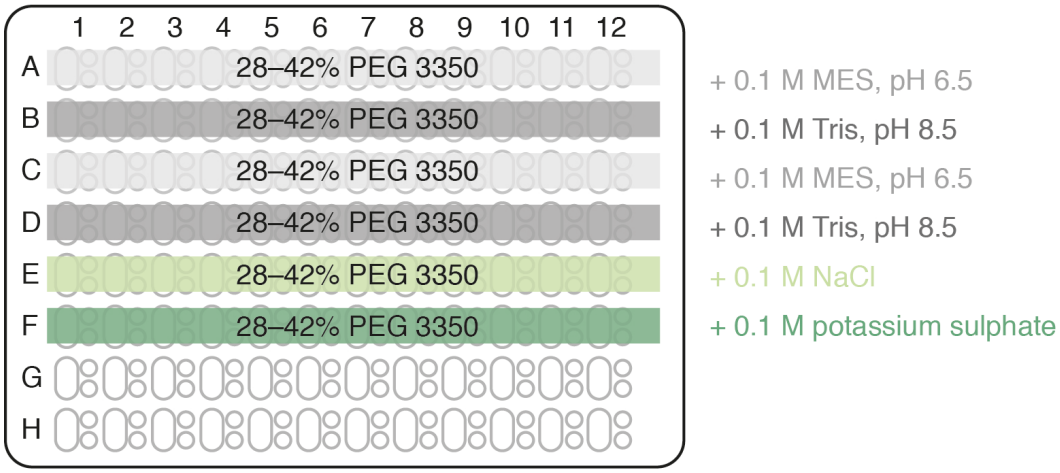
Appendix I

Supplementary information for Chapter 2: Materials and Methods

ancHMA$_{LVKIE}$/AVR-PikD



ancHMA$_{EMVKE}$/AVR-PikD (A–B) and ancHMA/AVR-PikD (C–F)



ancHMA/AVR-PikD

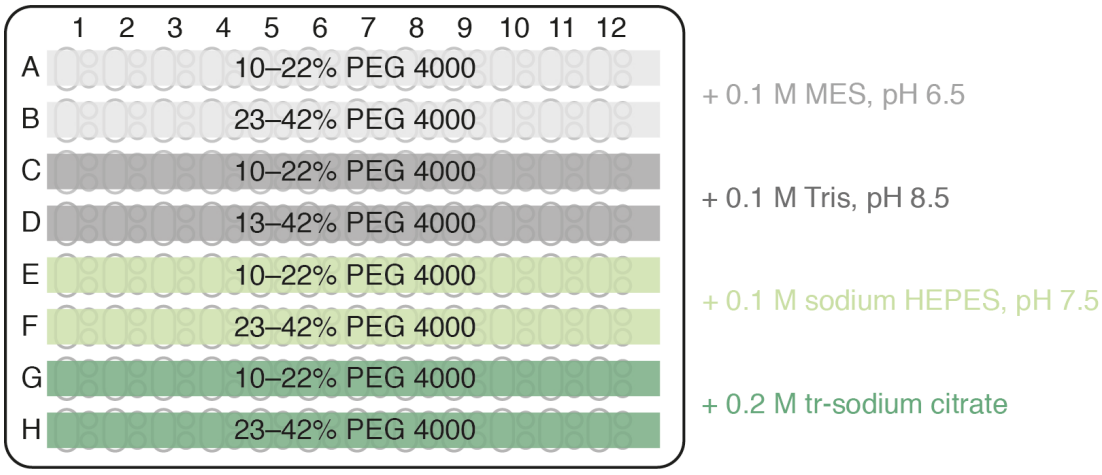


Figure A.1.1 Schematic illustration of optimisation of crystallisation conditions

Table A.1.1 Screen conditions of the JIC Custom Screen KISS (keep it simple screen)

Row	Column	Condition
A	1	0.1 M Sodium Acetate pH 4; 10% PEG 3350
A	2	0.1 M Sodium Acetate pH 4; 20% PEG 3350
A	3	0.1 M Sodium Acetate pH 4; 30% PEG 3350
A	4	0.1 M Sodium Acetate pH 4; 40% PEG 3350
A	5	0.1 M Sodium Citrate pH 5; 10% PEG 3350
A	6	0.1 M Sodium Citrate pH 5; 20% PEG 3350
A	7	0.1 M Sodium Citrate pH 5; 30% PEG 3350
A	8	0.1 M Sodium Citrate pH 5; 40% PEG 3350
A	9	0.1 M MES pH 6; 10% PEG 3350
A	10	0.1 M MES pH 6; 10% PEG 3350
A	11	0.1 M MES pH 6; 10% PEG 3350
A	12	0.1 M MES pH 6; 10% PEG 3350
B	1	0.1 M HEPES pH 7; 10% PEG 3350
B	2	0.1 M HEPES pH 7; 10% PEG 3350
B	3	0.1 M HEPES pH 7; 10% PEG 3350
B	4	0.1 M HEPES pH 7; 10% PEG 3350
B	5	0.1 M Tris pH 8; 10% PEG 3350
B	6	0.1 M Tris pH 8; 10% PEG 3350
B	7	0.1 M Tris pH 8; 10% PEG 3350
B	8	0.1 M Tris pH 8; 10% PEG 3350
B	9	0.1 M CHES pH 9; 10% PEG 3350
B	10	0.1 M CHES pH 9; 10% PEG 3350
B	11	0.1 M CHES pH 9; 10% PEG 3350
B	12	0.1 M CHES pH 9; 10% PEG 3350
C	1	0.8 M Ammonium Sulfate; 0.1 M Sodium Acetate pH 4
C	2	1.6 M Ammonium Sulfate; 0.1 M Sodium Acetate pH 4
C	3	2.4 M Ammonium Sulfate; 0.1 M Sodium Acetate pH 4
C	4	3.2 M Ammonium Sulfate; 0.1 M Sodium Acetate pH 4
C	5	0.8 M Ammonium Sulfate; 0.1 M Sodium Citrate pH 5
C	6	1.6 M Ammonium Sulfate; 0.1 M Sodium Citrate pH 5
C	7	2.4 M Ammonium Sulfate; 0.1 M Sodium Citrate pH 5
C	8	3.2 M Ammonium Sulfate; 0.1 M Sodium Citrate pH 5
C	9	0.8 M Ammonium Sulfate; 0.1 M MES pH 6
C	10	1.6 M Ammonium Sulfate; 0.1 M MES pH 6
C	11	2.4 M Ammonium Sulfate; 0.1 M MES pH 6
C	12	3.2 M Ammonium Sulfate; 0.1 M MES pH 6
D	1	0.8 M Ammonium Sulfate; 0.1 M HEPES pH 7
D	2	1.6 M Ammonium Sulfate; 0.1 M HEPES pH 7
D	3	2.4 M Ammonium Sulfate; 0.1 M HEPES pH 7
D	4	3.2 M Ammonium Sulfate; 0.1 M HEPES pH 7
D	5	0.8 M Ammonium Sulfate; 0.1 M Tris pH 8
D	6	1.6 M Ammonium Sulfate; 0.1 M Tris pH 8
D	7	2.4 M Ammonium Sulfate; 0.1 M Tris pH 8
D	8	3.2 M Ammonium Sulfate; 0.1 M Tris pH 8

Row	Column	Condition
D	9	0.8 M Ammonium Sulfate; 0.1 M CHES pH 9
D	10	1.6 M Ammonium Sulfate; 0.1 M CHES pH 9
D	11	2.4 M Ammonium Sulfate; 0.1 M CHES pH 9
D	12	3.2 M Ammonium Sulfate; 0.1 M CHES pH 9
E	1	0.2 M Ammonium Sulfate; 0.1 M Sodium Acetate pH 4; 10% PEG 3350
E	2	0.2 M Ammonium Sulfate; 0.1 M Sodium Acetate pH 4; 20% PEG 3350
E	3	0.2 M Ammonium Sulfate; 0.1 M Sodium Acetate pH 4; 30% PEG 3350
E	4	0.2 M Ammonium Sulfate; 0.1 M Sodium Acetate pH 4; 40% PEG 3350
E	5	0.2 M Ammonium Sulfate; 0.1 M Sodium Citrate pH 5; 10% PEG 3350
E	6	0.2 M Ammonium Sulfate; 0.1 M Sodium Citrate pH 5; 20% PEG 3350
E	7	0.2 M Ammonium Sulfate; 0.1 M Sodium Citrate pH 5; 30% PEG 3350
E	8	0.2 M Ammonium Sulfate; 0.1 M Sodium Citrate pH 5; 35% PEG 3350
E	9	0.2 M Ammonium Sulfate; 0.1 M MES pH 6; 10% PEG 3350
E	10	0.2 M Ammonium Sulfate; 0.1 M MES pH 6; 20% PEG 3350
E	11	0.2 M Ammonium Sulfate; 0.1 M MES pH 6; 30% PEG 3350
E	12	0.2 M Ammonium Sulfate; 0.1 M MES pH 6; 40% PEG 3350
F	1	0.2 M Ammonium Sulfate; 0.1 M HEPES pH 7; 10% PEG 3350
F	2	0.2 M Ammonium Sulfate; 0.1 M HEPES pH 7; 20% PEG 3350
F	3	0.2 M Ammonium Sulfate; 0.1 M HEPES pH 7; 30% PEG 3350
F	4	0.2 M Ammonium Sulfate; 0.1 M HEPES pH 7; 40% PEG 3350
F	5	0.2 M Ammonium Sulfate; 0.1 M Tris pH 8; 10% PEG 3350
F	6	0.2 M Ammonium Sulfate; 0.1 M Tris pH 8; 20% PEG 3350
F	7	0.2 M Ammonium Sulfate; 0.1 M Tris pH 8; 30% PEG 3350
F	8	0.2 M Ammonium Sulfate; 0.1 M Tris pH 8; 40% PEG 3350
F	9	0.2 M Ammonium Sulfate; 0.1 M CHES pH 9; 10% PEG 3350
F	10	0.2 M Ammonium Sulfate; 0.1 M CHES pH 9; 20% PEG 3350
F	11	0.2 M Ammonium Sulfate; 0.1 M CHES pH 9; 30% PEG 3350
F	12	0.2 M Ammonium Sulfate; 0.1 M CHES pH 9; 40% PEG 3350
G	1	0.5 M Ammonium Sulfate; 0.1M Sodium Acetate pH 4; 5% PEG 3350
G	2	0.6 M Ammonium Sulfate; 0.1M Sodium Acetate pH 4; 5% PEG 3350
G	3	0.8 M Ammonium Sulfate; 0.1M Sodium Acetate pH 4; 5% PEG 3350
G	4	1.0 M Ammonium Sulfate; 0.1M Sodium Acetate pH 4; 5% PEG 3350
G	5	0.5 M Ammonium Sulfate; 0.1M Sodium Citrate pH 5; 5% PEG 3350
G	6	0.6 M Ammonium Sulfate; 0.1M Sodium Citrate pH 5; 5% PEG 3350
G	7	0.8 M Ammonium Sulfate; 0.1M Sodium Citrate pH 5; 5% PEG 3350
G	8	1.0 M Ammonium Sulfate; 0.1M Sodium Citrate pH 5; 5% PEG 3350
G	9	0.5 M Ammonium Sulfate; 0.1M MES pH 6; 5% PEG 3350
G	10	0.6 M Ammonium Sulfate; 0.1M MES pH 6; 5% PEG 3350
G	11	0.8 M Ammonium Sulfate; 0.1M MES pH 6; 5% PEG 3350
G	12	1.0 M Ammonium Sulfate; 0.1M MES pH 6; 5% PEG 3350
H	1	0.5 M Ammonium Sulfate; 0.1M HEPES pH 7; 5% PEG 3350
H	2	0.6 M Ammonium Sulfate; 0.1M HEPES pH 7; 5% PEG 3350
H	3	0.8 M Ammonium Sulfate; 0.1M HEPES pH 7; 5% PEG 3350
H	4	1.0 M Ammonium Sulfate; 0.1M HEPES pH 7; 5% PEG 3350

Row	Column	Condition
H	5	0.5 M Ammonium Sulfate; 0.1M Tris pH 8; 5% PEG 3350
H	6	0.6 M Ammonium Sulfate; 0.1M Tris pH 8; 5% PEG 3350
H	7	0.8 M Ammonium Sulfate; 0.1M Tris pH 8; 5% PEG 3350
H	8	1.0 M Ammonium Sulfate; 0.1M Tris pH 8; 5% PEG 3350
H	9	0.5 M Ammonium Sulfate; 0.1M CHES pH 9; 5% PEG 3350
H	10	0.6 M Ammonium Sulfate; 0.1M CHES pH 9; 5% PEG 3350
H	11	0.8 M Ammonium Sulfate; 0.1M CHES pH 9; 5% PEG 3350
H	12	1.0 M Ammonium Sulfate; 0.1M CHES pH 9; 5% PEG 3350

Appendix II

Supplementary information for Chapter 3: The HMA integration into Pik-1 predates the emergence of Oryzinae

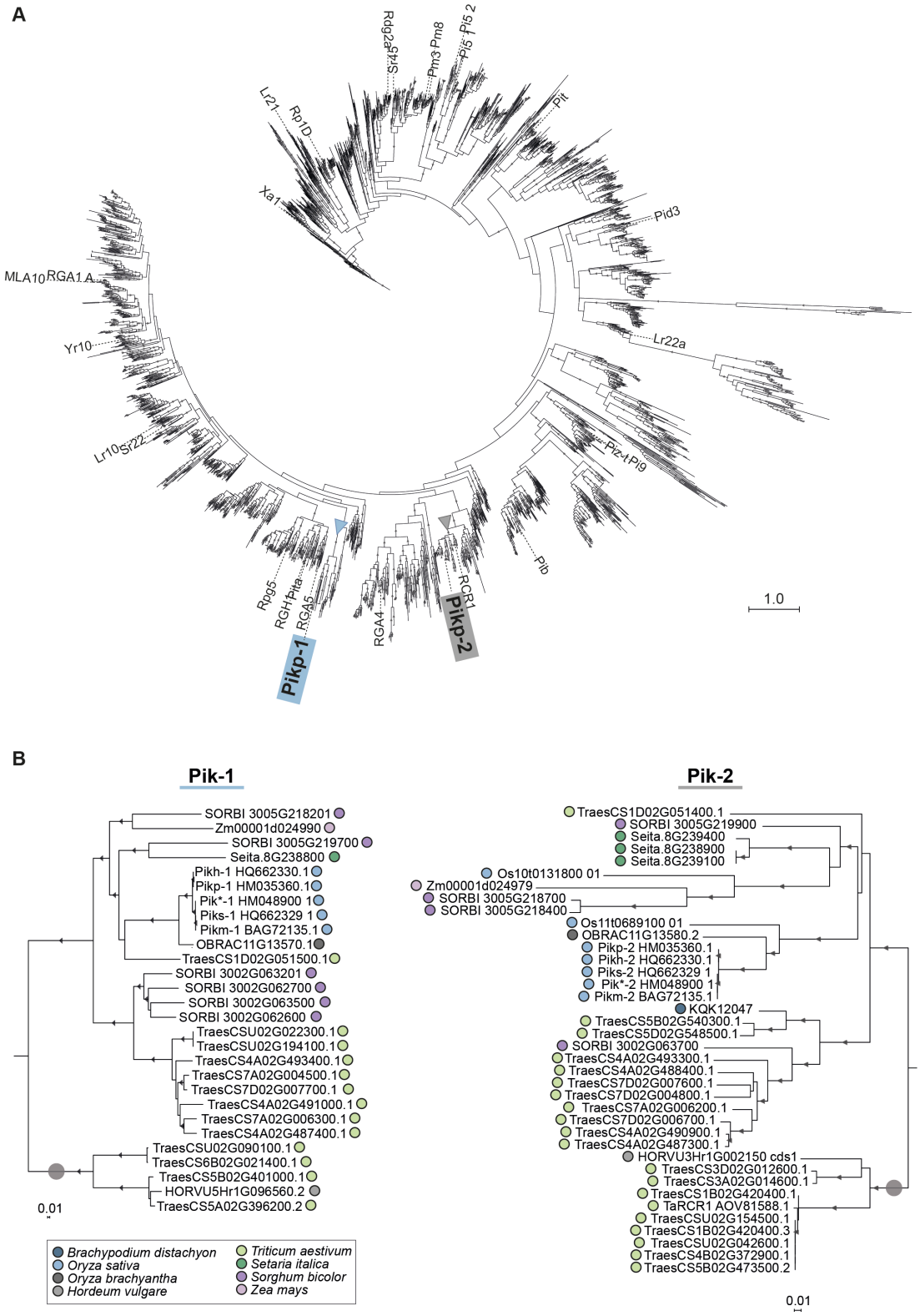


Figure A.2.1 Pik-1 and Pik-2 orthologues fall into two well-supported clades

(**A**) Phylogenetic tree of CC-type NLRs of *Zea mays*, *Sorghum bicolor*, *Setaria italica*, *Triticum aestivum*, *Hordeum vulgare*, *Brachypodium distachyon*, *Oryza brachyantha*, and *Oryza sativa*. The maximum

likelihood (ML) tree was calculated based on 241-amino-acid-long alignment of 3,062 CC-NLRs amended with 35 known and functionally characterized NLRs from grasses using RAxML v8.2.11 (Stamatakis, 2014) with bootstrap values (Felsenstein, 1985) based on 1000 iterations and using the best-scoring JTT likelihood model (Jones et al., 1992). The best ML tree is shown. The scale bar indicates the evolutionary distance based on site substitution rate. The clades constituting Pik-1 and Pik-2 homologues are marked with blue and grey triangles, respectively. Branches corresponding to the reference NLRs are labelled. The interactive tree is publicly available at: https://itol.embl.de/tree/822913314736537160286345 7. (**B**) The ML phylogenetic trees of Pik-1 (left) and Pik-2 (right) homologues constructed based on 957- and 1218-nucleotides-long codon-based alignments, respectively, using RAxML v8.2.11 (Stamatakis, 2014), 1000 bootstrap method (Felsenstein, 1985), and GTRGAMMA substitution model (Tavaré, 1986). Best ML trees were manually rooted based on previously observed relationships, where clades marked with grey circle were used as outgroups. Bootstrap values above 70% are marked with grey triangles at the base of respective clades. The scale bars indicate the evolutionary distance based on nucleotide substitution rate. The interactive trees are publicly available at: https://itol.embl.de/tree/822913314744949160286481 2
and https://itol.embl.de/tree/822913314744951160286481 2.

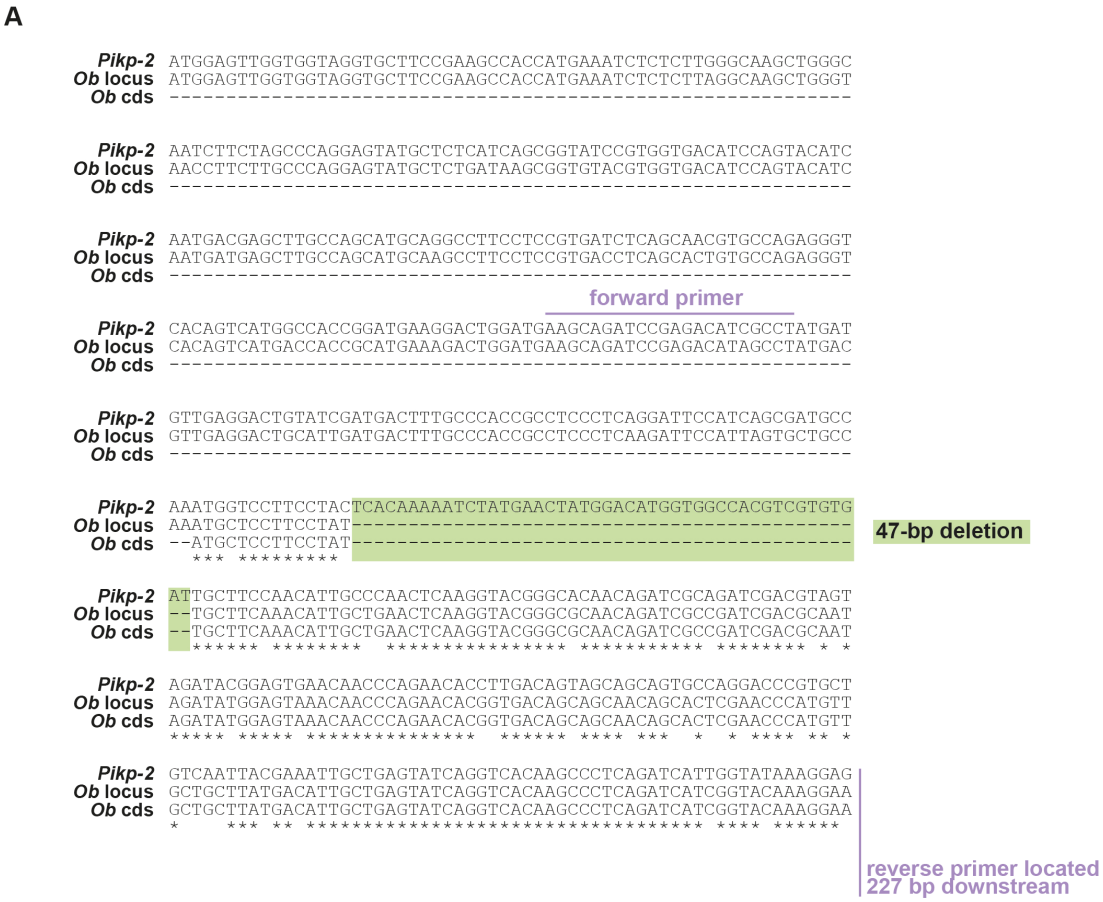


Figure A.2.2 Genotyping of *Oryza brachyantha* accessions

(**A**) Nucleotide alignment of *Pikp-2*, the *ObPik-2* (*Ob* locus) gene, and the *ObPik-2* coding sequence (*Ob* cds) from the reference genome (Chen et al., 2013), illustrating 47-bp-long deletion and the primers used for the genotyping. (**B**) Gel electrophoresis of *Ob*Pik-2 fragments amplified from different *Oryza brachyantha* accessions, labelled above. The symbols next to the accession numbers mark sequences that: carry the 47-bp deletion (*), harbour 4-bp deletion (**), don't carry any deletions (•), were used for amplification of the full-length gene (#). Water and *Pikp-1* were used as negative and positive controls, respectively. The left and the right lanes show molecular size markers.

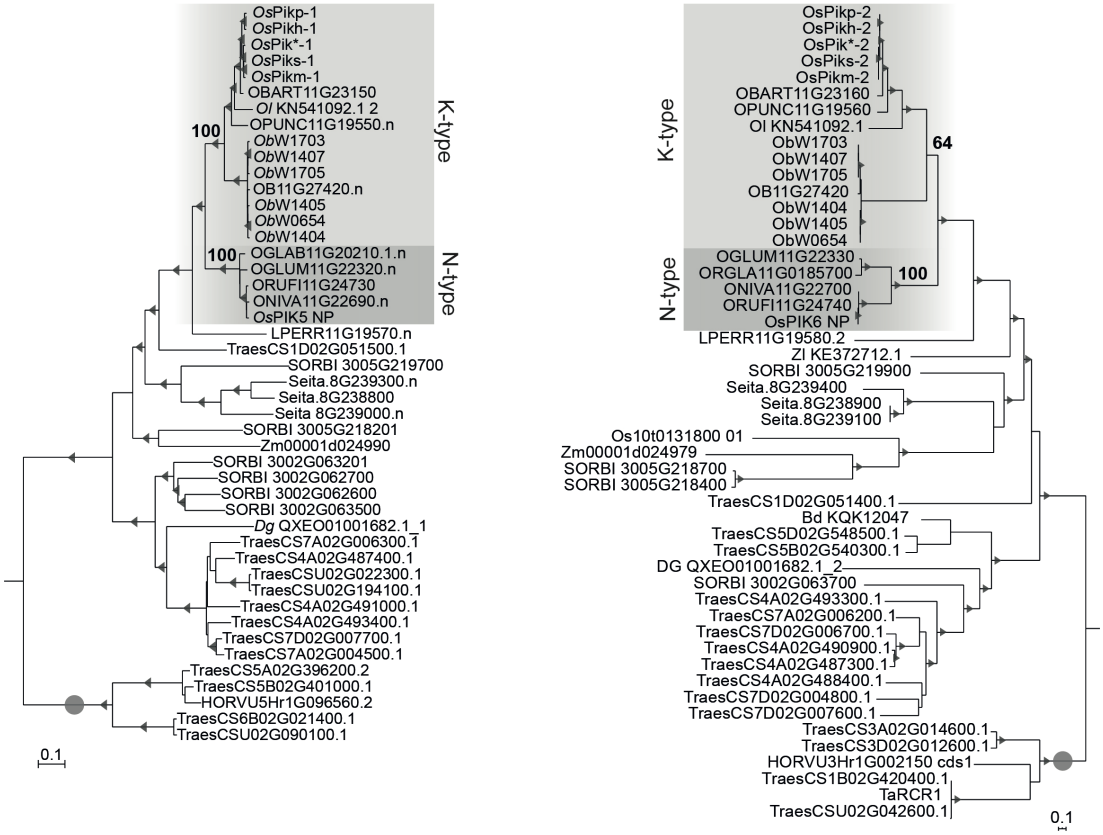


Figure A.2.3 Pik-1 and Pik-2 orthologues from *Oryza* fall into K- and N-type clades

The phylogenetic tree shown in **Figure 3.1 A**, illustrating the divide between the N- (dark grey) and K-type (light grey) *Pik* genes. The bootstrap values for the relevant clades are marked.

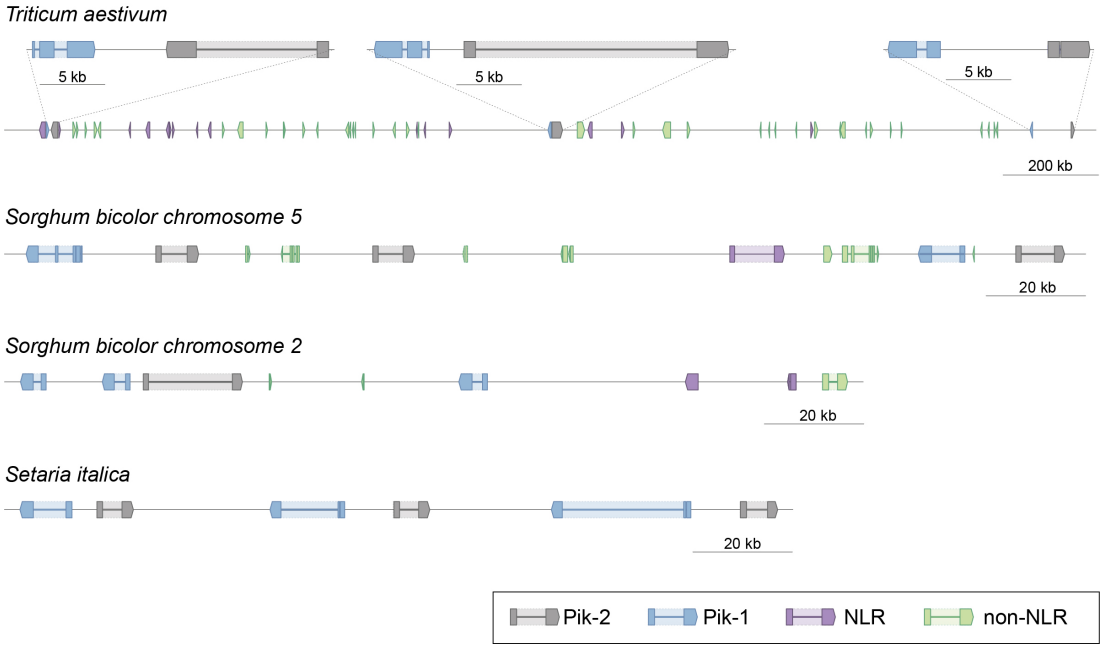


Figure A.2.4 Schematic representation of selected *Pik* clusters in wheat (*T. aestivum*), sorghum (*S. bicolor*), and foxtail millet (*S. italica*)

The schematic presents gene models and genetic locations of *Pik-1* (blue), *Pik-2* (grey), and other NLR genes (purple). Non-NLR genes are shown in light green. The coordinates of the regions presented in this figure are summarised in **Table A.2.2**.

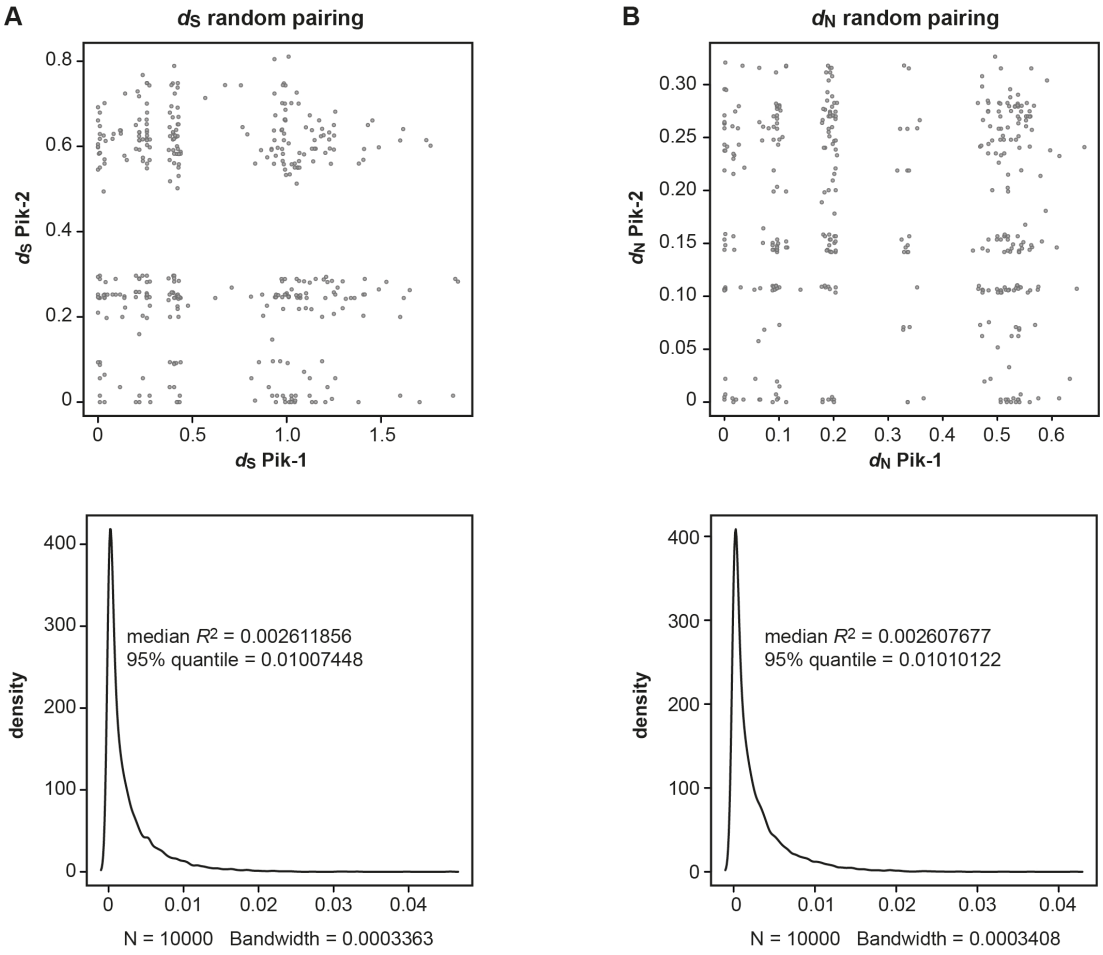


Figure A.2.5 Random pairwise comparisons of d_S (**A**) and d_N (**B**) rates calculated for the Pik-1 and Pik-2 receptors

The synonymous (d_S) and nonsynonymous (d_N) rates were calculated using Yang and Nielsen method (Yang and Nielsen, 2000) using the dataset summarised in **Table A.2.3**. The random datasets for d_S and d_N values were generated by name shuffling in the existing dataset and random sampling from it 1000 times (upper panels). The coefficient of determination (R^2) was calculated for every random pairing and the R^2 distribution was plotted (lower panels), as implemented in R v3.6.3 package. If less than 5% of the R^2 for the random dataset is bigger than the R^2 for the real dataset, then, according to the null model, the observed difference is very rare and can be therefore accepted as significant with $p < 0.05$.

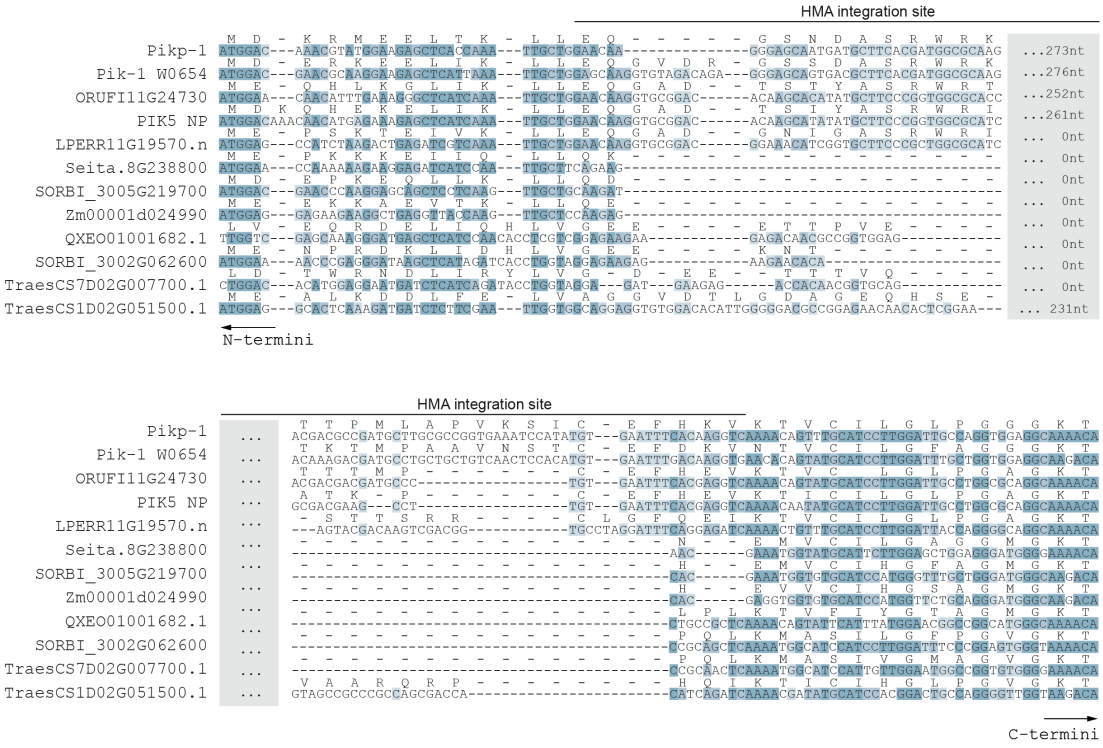


Figure A.2.6 The multiple sequence alignment illustrating conservation around the HMA integration site

The codon-based sequence alignemed of the region surrounding the HMA integration site was generated using MUSCLE (Edgar, 2004) and visualised using JalView (Waterhouse et al., 2009). The residues are coloured based on percentage identity from dark (high identity) to light (low identity) blue.

Model	Estimates of Parameters	lnL
M0 (one-ratio)	$\omega = 0.48555$	-2407.8465
M1 (nearly neutral)	$P_0 = 0.48633$, $P_1 = 0.51367$	-2395.5229
M2 (selection)	$P_0 = 0.48599$, $P_1 = 0.50364$, $P_2 = 0.01037$ $\omega_0 = 0.00000$, $\omega_1 = 1.00000$, $\omega_2 = 2.70133$	-2395.4752
M3 (discrete)	$P_0 = 0.49014$, $P_1 = 0.50440$, $P_2 = 0.00546$ $\omega_0 = 0.00000$, $\omega_1 = 1.02509$, $\omega_2 = 3.11743$	-2395.4720
M7 (beta)	$P_0 = 0.01350$, q = 0.01216	-2395.5555
M8 (beta + ω)	$P_0 = 0.49595$, $P = 0.00500$, $q = 1.95516$ ($P_1 = 0.50405$), $\omega = 1.05554$	-2395.4835

Model comparison	2ΔL	χ^2 (0.001)	df
M1 vs. M2	0.095558	13.816	2
M0 vs. M3	24.748954	18.467	4
M7 vs. M8	0.072017	13.816	2

Figure A.2.7 Selection test at the amino acid sites within the NB-ARC domain of the K-type Pik-1 orthologues

Results from the codon substitution models for heterogeneous selection at amino acid sites (upper panel) and the likelihood ratio test (bottom panel).

Table A.2.1 Coordinates of genomic regions used in **Figure 3.2**

Name	Species	Contig/ Chromosome	Start	End
Os	*Oryza sativa* cv. K60	Chromosome 11	27973977	28008157
Oniv	*Oryza nivara*	Chromosome 11	23864010	23917629
Oglum	*Oryza glumaepatula*	Chromosome 11	26116594	26116594
Ol	*Oryza longistaminata*	Contig CM003669.1	2965866	8003128
Opunc	*Oryza punctata*	Chromosome 11	4972847	4983002
Ob	*Oryza brachyantha*	Chromosome 11	15529280	15547390
Lp	*Leersia perrieri*	Chromosome 11	20337647	20286182
Ta	*Triticum aestivum*	Chromosome 1D	33124348	31125148
Dg	*Dactylis glomerats*	Scaffold QXEO01001682.1	1295679	1340805
Si	*Setaria italica*	Scaffold 8	39159743	39261506
Sb	*Sorghum bicolor*	Chromosome 2	6043453	6215456

Table A.2.2 Coordinates of genomic regions used in **Figure A.2.4**

Species	Chromosome / Scaffold	Start	End
Triticum aestivum	Chromosome 1D	33124348	31125148
Triticum aestivum	Chromosome 4A	739372651	742475497
Triticum aestivum	Chromosome 7D	3066514	4338541
Sorghum bicolor	Chromosome 5	70290712	70702146
Sorghum bicolor	Chromosome 2	5911252	6248983
Setaria italica	Scaffold 8	39100454	39349775

Table A.2.3 Genes used for the comparisons of d_S and d_N rates of Pik-1–Pik-2 presented in **Figure 3.3**

Name	Pik-1	Pik-2	Species
***Os*Pikp**	Pikp1_HM035360.1	Pikp2_HM035360.1	*Oryza sativa*
***Os*Pikh**	Pikh_1_HQ662330.1	Pikh2_HQ662330.1	*Oryza sativa*
Os*Pik	Pik_1_HM048900_1	Pik2_HM048900_1	*Oryza sativa*
***Os*Piks**	Piks_1_HQ662329_1	Piks2_HQ662329_1	*Oryza sativa*
***Os*Pikm**	Pikm_1_BAG72135.1	Pikm2_BAG72135.1	*Oryza sativa*
Obart Pik	OBART11G23150	OBART11G23160	*Oryza barthii*
Olongi Pik	KN541092.1_2	KN541092.1	*Oryza longistaminata*
Opunc Pik	OPUNC11G19550.n	OPUNC11G19560	*Oryza punctata*
***Ob*Pik W1703**	ObPik_1_W1703	ObPik2_W1703_CDS	*Oryza brachyantha*
***Ob*Pik W1407**	ObPik_1_W1407	ObPik2_W1407_CDS	*Oryza brachyantha*
***Ob*Pik W1705**	ObPik_1_W1705	ObPik2_W1705_CDS	*Oryza brachyantha*
***Ob*Pik IRGC101232**	OB11G27420.n	OB11G27420	*Oryza brachyantha*
***Ob*Pik W1405**	ObPik_1_W1405	ObPik2_W1405_CDS	*Oryza brachyantha*
***Ob*Pik W1404**	ObPik_1_W0654	ObPik2_W0654_CDS	*Oryza brachyantha*
***Ob*Pik W0654**	ObPik_1_W1404	ObPik2_W1404_CDS	*Oryza brachyantha*
Oglum Pik	OGLAB11G20210.1n	ORGLA11G0185700	*Oryza glaberrima*
Oglab Pik	OGLUM11G22320.n	OGLUM11G22330	*Oryza glumaepatula*
Oniv Pik	ORUFI11G24730	ORUFI11G24740	*Oryza rufipogon*
Oruf Pik	ONIVA11G22690.n	ONIVA11G22700	*Oryza nivara*
***Os*Pik Nipp**	Pikm5_NP_Nipp_DP000010.2	PIK6_NP_XM_015762499.2	*Oryza sativa* cv. Nipponbare
***Ta*Pik 1D**	TraesCS1D02G051500.1	TraesCS1D02G051400.1	*Triticum aestivum*
***Sb*Pik 5**	SORBI_3005G219700	SORBI_3005G219900	*Sorghum bicolor*
***Si*Pik 8.1**	Seita.8G239300.n	Seita.8G239400	*Setaria italica*
***Si*Pik 8.2**	Seita.8G238800	Seita.8G238900	*Setaria italica*
***Dg*Pik**	QXEO01001682.1	QXEO01001682.1_2	*Dactylis glomerata*
***Ta*Pik 4A.1**	TraesCS4A02G493400.1	TraesCS7A02G006200.1	*Triticum aestivum*
***Ta*Pik 4A.2**	TraesCS4A02G491000.1	TraesCS4A02G490900.1	*Triticum aestivum*
***Ta*Pik 7D**	TraesCS7D02G007700.1	TraesCS7D02G007600.1	*Triticum aestivum*

Table A.2.4 Summary of the amplification experiment of the Pik-1–integrated HMA domain from wild rice species

Accession	Species	Origin	Amplified	Sequence confirmed
W0654	*O. brachyantha*	Sierra Leone	Yes	Yes
W0008	*O. australiensis*	Australia (SE Canberra)	Yes	No
W1628	*O. australiensis*	Australia (N)	No	NA
W1643	*O. barthii*	Botswana	Yes	Yes
W1605	*O. barthii*	Nigeria	Yes	No
W0042	*O. barthii*	unspecified	Yes	Yes
W0698	*O. barthii*	Guinea	Yes	Yes
W1526	*O. eichingeri*	Uganda	No	NA
W1171	*O. glumaepatula*	Cuba	No	NA
W2203	*O. glumaepatula*	Brasil (S)	Yes	No
W1480(B)	*O. grandiglumis*	Brasil (N)	Yes	No
W0005	*O. granulata*	Sri Lanka	No	NA
W0067(B)	*O. granulata*	Thailand	Yes	Yes
W0542	*O. latifolia / O. alta*	Mexico	No	NA
W1539	*O. latifolia / O. alta*	Argentina (N)	No	NA
W1228	*O. longiglumis*	Singapore (S)	No	NA
W1504	*O. longistaminata*	Tanzania	No	NA
W1540	*O. longistaminata*	Republic of Congo	Yes	Yes
W0643	*O. longistaminata*	The Gambia	Yes	Yes
W2081	*O. meridionalis*	Australia (N)	No	NA
W2112	*O. meridionalis*	Australia (NE)	No	NA
W1354	*O. meyeriana*	Malaysia	Yes	No
W1328	*O. minuta*	Philippines	Yes	Yes
W0614	*O. officinalis*	Myanmar	Yes	Yes
W1200	*O. officinalis*	Philippines	Yes	Yes
W1408	*O. punctata*	Nigeria	Yes	Yes
W1514	*O. punctata*	Kenya	Yes	Yes
W1808	*O. rhizomatis*	Sri Lanka	Yes	No
W0001	*O. ridleyi*	Thailand	Yes	No
W2035	*O. ridleyi*	Philippines	No	NA
W2003	*O. rufipogon*	India (SW)	Yes	Yes
W1715	*O. rufipogon*	Chin (Beijing)	No	NA
W2117	*O. rufipogon* / *O. meridionalis*	Australia (NE)	Yes	No

Appendix III

Supplementary information for Chapter 4: Pikp-1 evolved towards high affinity binding to the AVR-PikD effector from the rice blast fungus

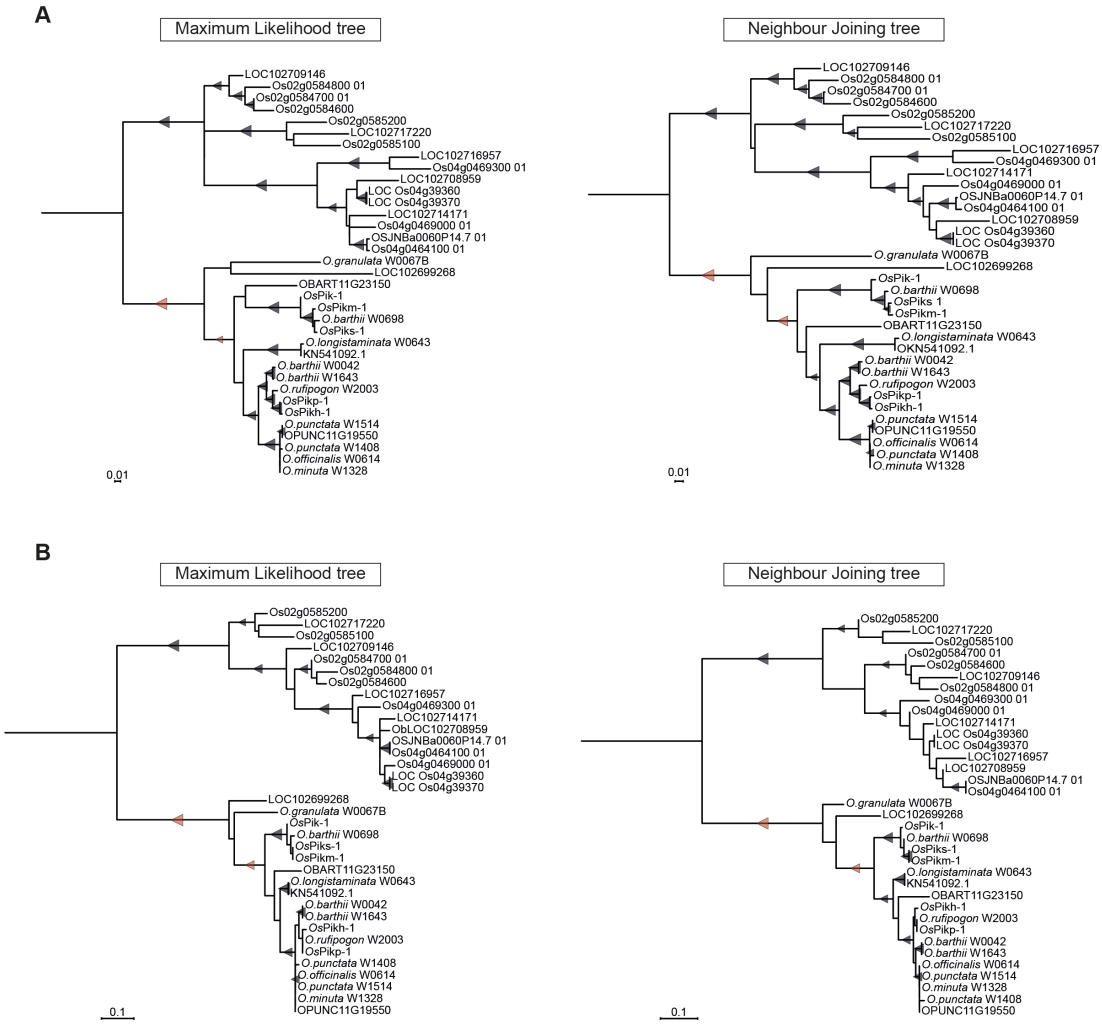


Figure A.3.1 Phylogenetic analyses of the HMA domain of the K-type Pik-1 NLRs

The phylogenetic trees were built using MEGA X software (Kumar et al., 2018) and bootstrap method based on 1000 iterations (Felsenstein, 1985). Codon-based 249-nucleotide-long alignment was generated using MUSCLE (Edgar, 2004); positions with less than 50% site coverage were removed prior the analysis, resulting in a total of 234 positions in the final dataset. The relevant bootstrap values with support over 60% are shown with triangles at the base of representative clades; the size of the triangle is proportional to the bootstrap value. The scale bars indicate the evolutionary distance based on nucleotide substitution rate. Each tree was manually rooted using a clade of non-integrated HMA as an outgroup. The nodes selected for ancestral sequence reconstruction are marked with red triangles. (**A**) Maximum likelihood and neighbour joining trees calculated based on all codon positions in the alignment. (**B**) Maximum likelihood and neighbour joining trees calculated based on a third codon position in the alignment.

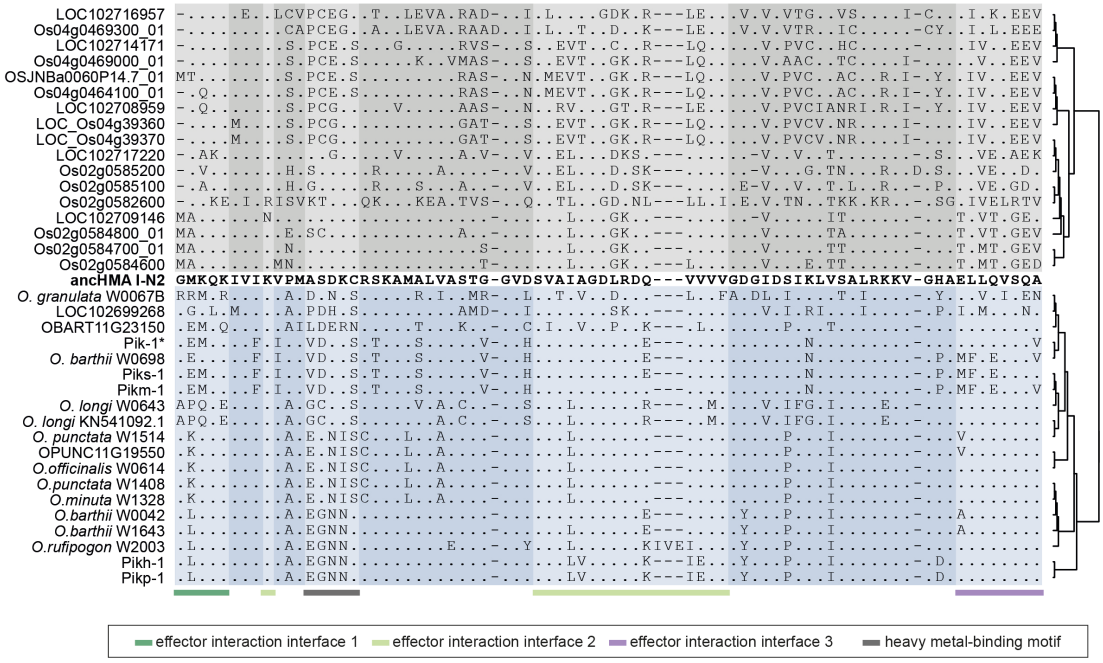


Figure A.3.2 Protein sequence alignment of the HMA domain

Sequences of the K-type Pik-1–integrated HMA domains (blue), non-integrated HMAs from *Oryza sativa* and *Oryza brachyantha* (grey), and I-N2 ancHMA (bold) were aligned using MUSCLE (Edgar, 2004). Regions with known function are marked with horizontal lines at the bottom.

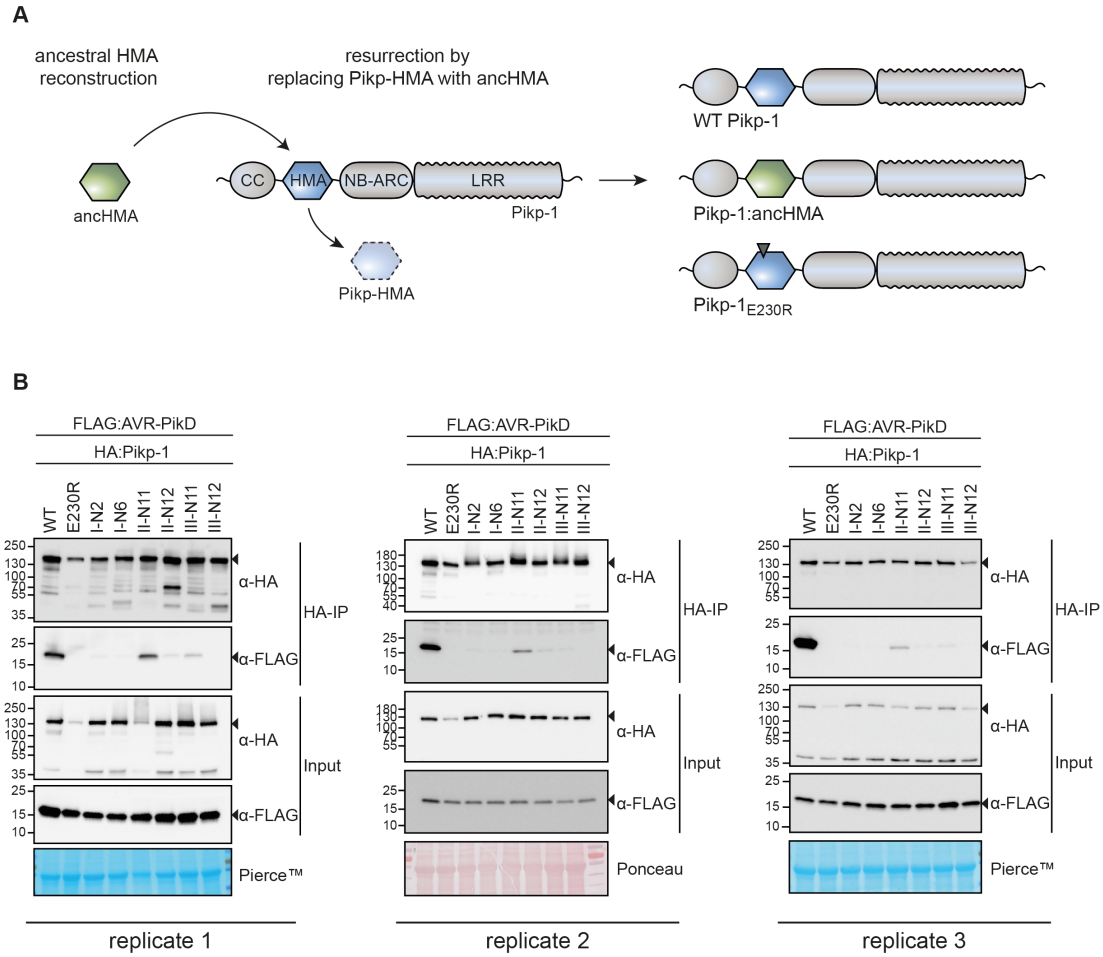


Figure A.3.3 The integrated HMA domain of Pikp-1 evolved towards strong binding to the AVR-PikD effector from a weaker ancestral state

(**A**) Overview of the strategy for resurrection of the ancestral HMA (ancHMA) domain. Following ancestral sequence reconstruction, the appropriate gene sequences were synthetized and incorporated into Pikp-1 by replacing the present-day Pikp-HMA domain (blue) with the ancHMA equivalent (green). (**B**) Co-immunoprecipitation experiment between FLAG-tagged AVR-PikD with HA-tagged Pikp-1 consisting of ancestral sequences of the HMA—Pikp-1:I-N2, Pikp-1:I-N6, Pikp-1:II-N11, Pikp-1:II-N12, Pikp-1:III-N11, and Pikp-1:III-N12. Wild type (WT) HA:Pikp-1 and HA:Pikp-1$_{E230R}$ were used as positive and negative controls, respectively. Proteins obtained by co-immunoprecipitation with HA-probe (HA-IP) and total protein extracts (Input) were immunoblotted with appropriate antisera (listed on the right). Rubisco loading control was performed using the Pierce™ or Ponceau staining solutions. Arrowheads indicate expected band sizes. The figure features results from three independent experiments.

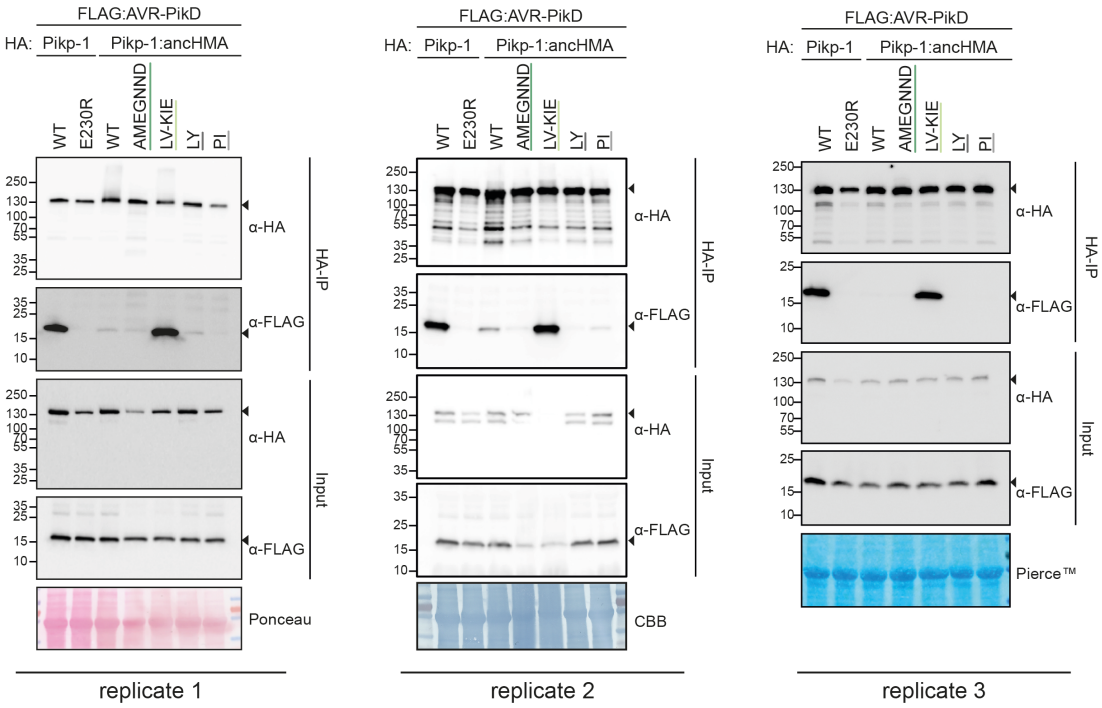


Figure A.3.4 Replicates of the co-IP experiment between AVR-PikD and the Pikp–ancHMA chimeras

Association of FLAG-tagged AVR-PikD with HA-tagged Pikp-1, Pikp-1_{E230R}, Pikp-1:ancHMA, and Pikp-1:ancHMA chimeras, labelled above, was tested in planta in co-immunoprecipitation (co-IP). Wild type (WT) Pikp-1 and Pikp-1_{E230R}, featuring HA tag, were used as positive and negative controls, respectively. Total protein extracts (Input) and proteins obtained by co-immunoprecipitation with HA-probe (HA-IP) were immunoblotted with the appropriate antisera labelled on the right. Arrowheads show expected band sizes. Rubisco loading controls were performed using Pierce™, Coomassie Brilliant Blue (CBB), or Ponceau staining solutions. The figure shows results from three independent experiments.

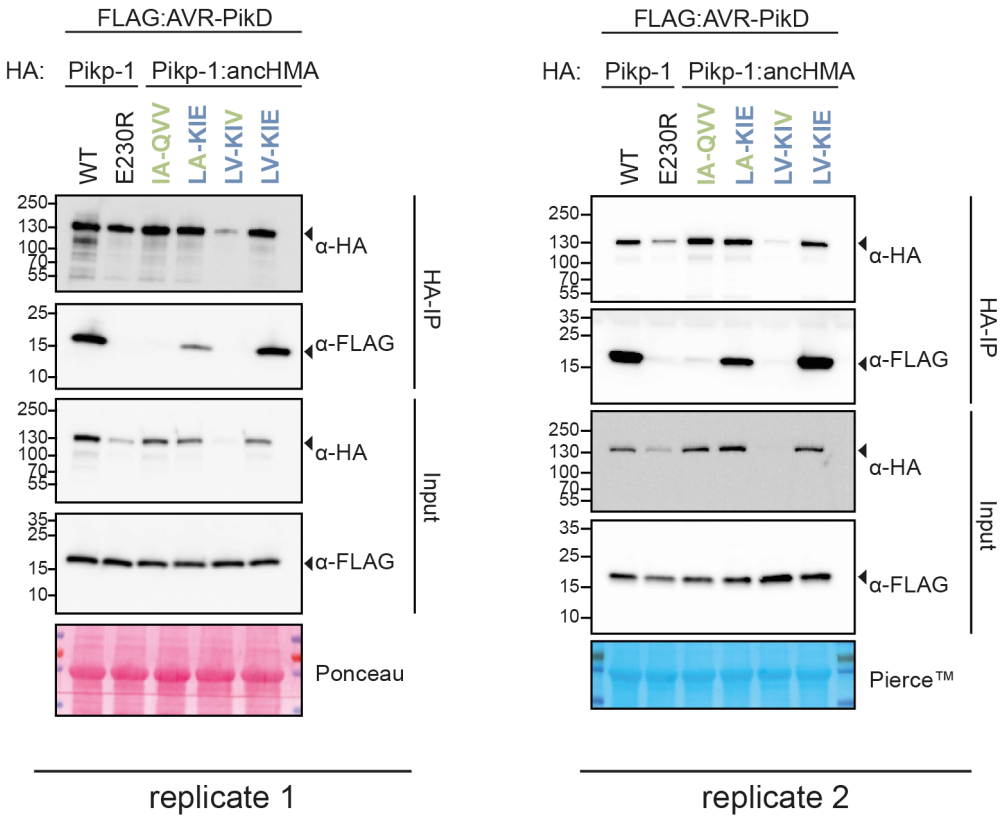


Figure A.3.5 Co-immunoprecipitation (co-IP) experiment between AVR-PikD and the two plausible historical states of the LVKIE region within Pikp-HMA

In planta association of FLAG-tagged AVR-PikD with HA-tagged Pikp-1, Pikp-1$_{E230R}$, Pikp-1:ancHMA, and Pikp-1:ancHMA mutants, labelled above. Wild type (WT) Pikp-1 and Pikp-1$_{E230R}$, with HA tag, were used as positive and negative controls, respectively. Proteins obtained by co-IP with HA resin (HA-IP) and total protein extracts (Input) were immunoblotted with the appropriate antisera labelled on the right. Arrowheads indicate expected band sizes. Loading controls, featuring rubisco, were performed using Pierce™ or Ponceau staining solutions. The figure features results from two independent experiments.

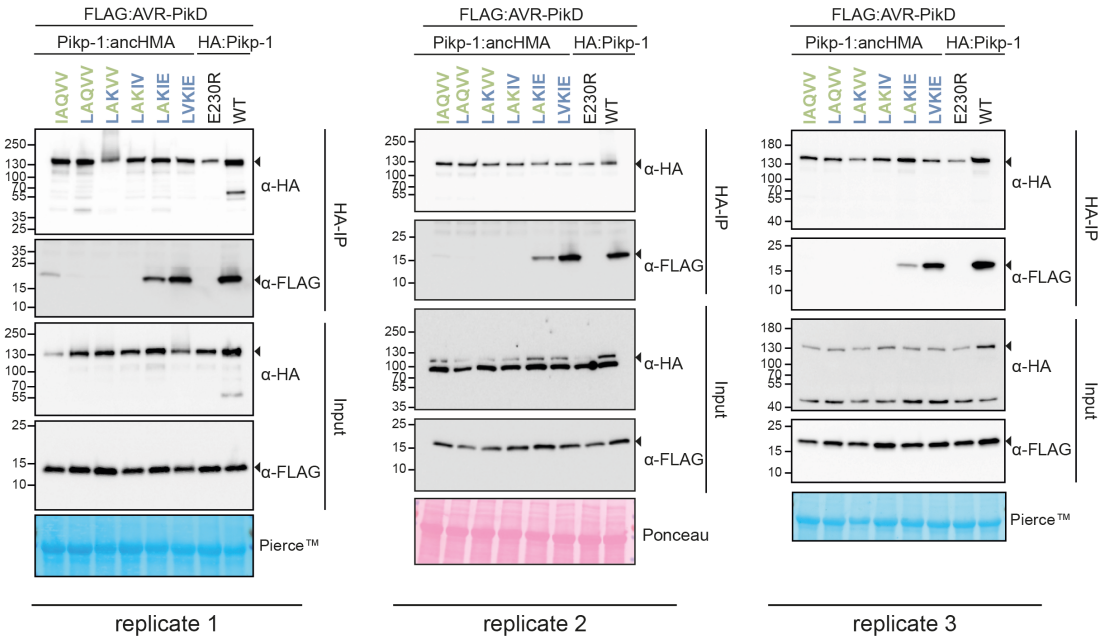
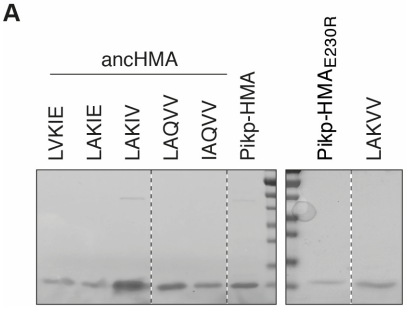


Figure A.3.6 Replicates of co-IP experiments between the Pikp-1:ancHMA mutants and AVR-PikD

In planta association of FLAG-tagged AVR-PikD with HA-tagged Pikp-1, Pikp-1$_{E230R}$, Pikp-1:ancHMA, and Pikp-1:ancHMA mutants, labelled above. Wild type (WT) Pikp-1 and Pikp-1$_{E230R}$, featuring HA tag, were used as positive and negative controls, respectively. Proteins obtained by co-immunoprecipitation with HA-probe (HA-IP) and total protein extracts (Input) were immunoblotted with the appropriate antisera labelled on the right. Rubisco loading controls were conducted using Pierce™ or Ponceau staining solutions. Arrowheads demonstrate expected band sizes. The figure shows the results from three biological experiments.

A



B

	Sample name	Predicted mass (Da)	Observed mass (Da)	Difference
	Pikp-HMA	7800.3199	7800.3395	-0.0196
	Pikp-HMA$_{E230R}$	8383.639	8383.595	0.044
ancHMA	LVKIE	7781.2923	7781.2897	0.0026
ancHMA	LAKIE	7753.261	7753.2567	0.0043
ancHMA	LAKIV	7723.2868	7723.2936	-0.0068
ancHMA	LAKVV	7714.19	7715.0005	0.8103
ancHMA	LAQVV	7709.2348	7709.1953	0.0395
ancHMA	IAQVV	7709.2348	7709.1889	0.0459

Figure A.3.7 Purified proteins used in surface plasmon resonance (SPR) studies

(**A**) Coomasie Brilliant Blue–stained SDS-PAGE gel showing purified HMA proteins used in in vitro experiments. Dashed line signifies different components of the same gel. (**B**) Table summarising intact masses of proteins from panel A, as measured by The Sainsbury Laboratory core Proteomics facilities (Norwich, UK).

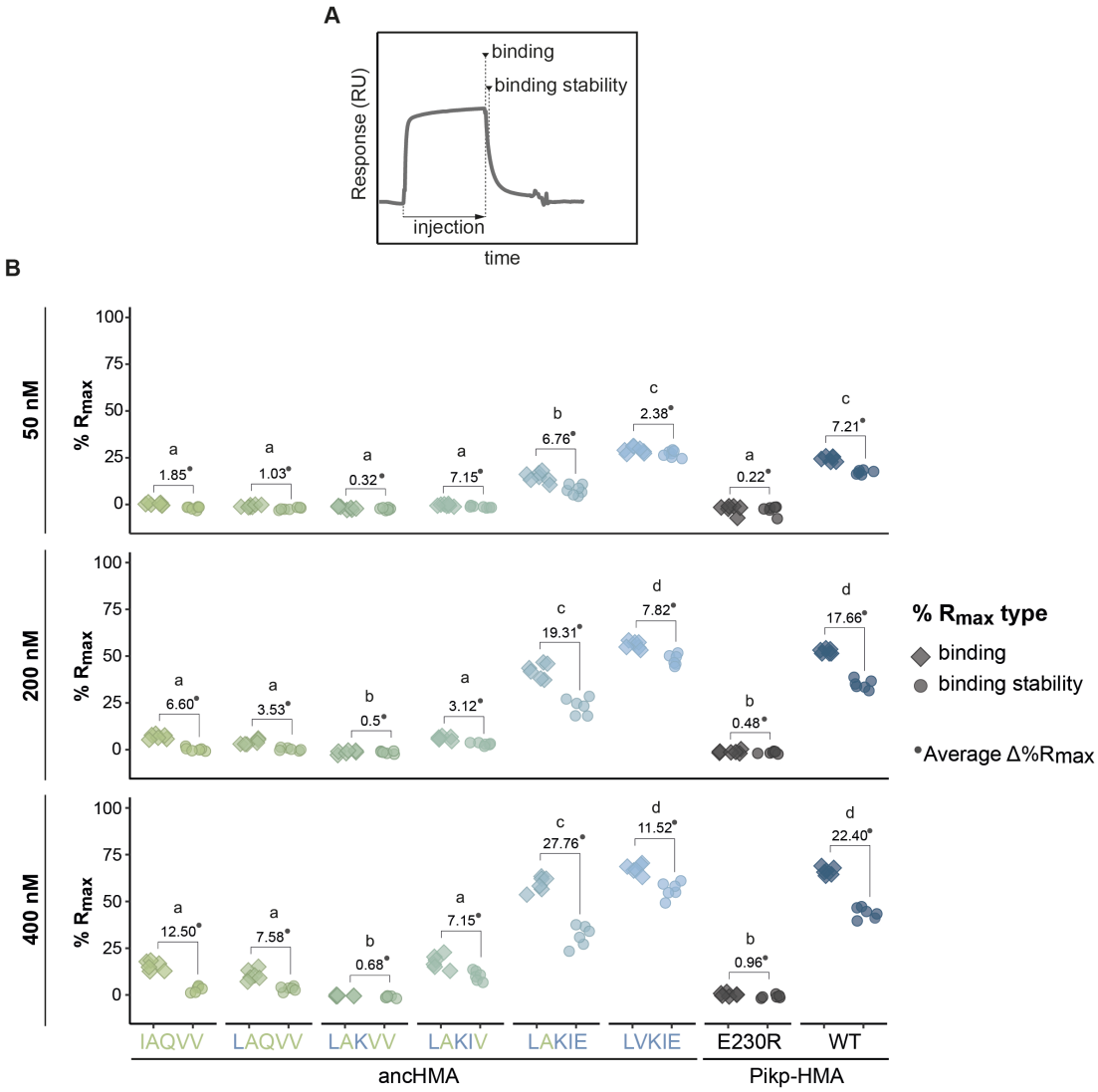


Figure A.3.8 Surface plasmon resonance (SPR) results show the effect of the LVKIE mutations on the AVR-PikD binding, as indicated by $\%R_{max}$

(**A**) Schematic representation of the SPR sensograms showcasing the measurements taken to monitor binding dynamics: 'binding' and 'binding stability'. (**B**) Plots illustrating calculated percentage of the theoretical maximum response values ($\%R_{max}$) for interaction of the HMA analytes, labelled below, with His-tagged AVR-PikD ligand. $\%R_{max}$ was normalized for the amount of ligand immobilized on the NTA-sensor chip. The HMA analytes were tested at three different concentrations, indicated on the left, in three independent experiments with two internal replicates. All data points are represented as diamonds or circles. Average $\Delta\%R_{max}$ (•) values represent absolute differences between values for 'binding' and 'biding stability' calculated from average values for each of the samples. Statistical differences among the samples were analysed with ANOVA and Tukey's honest significant difference (HSD) test ($p < 0.01$). P-values for all pairwise comparisons are presented in **Table A.3.2**.

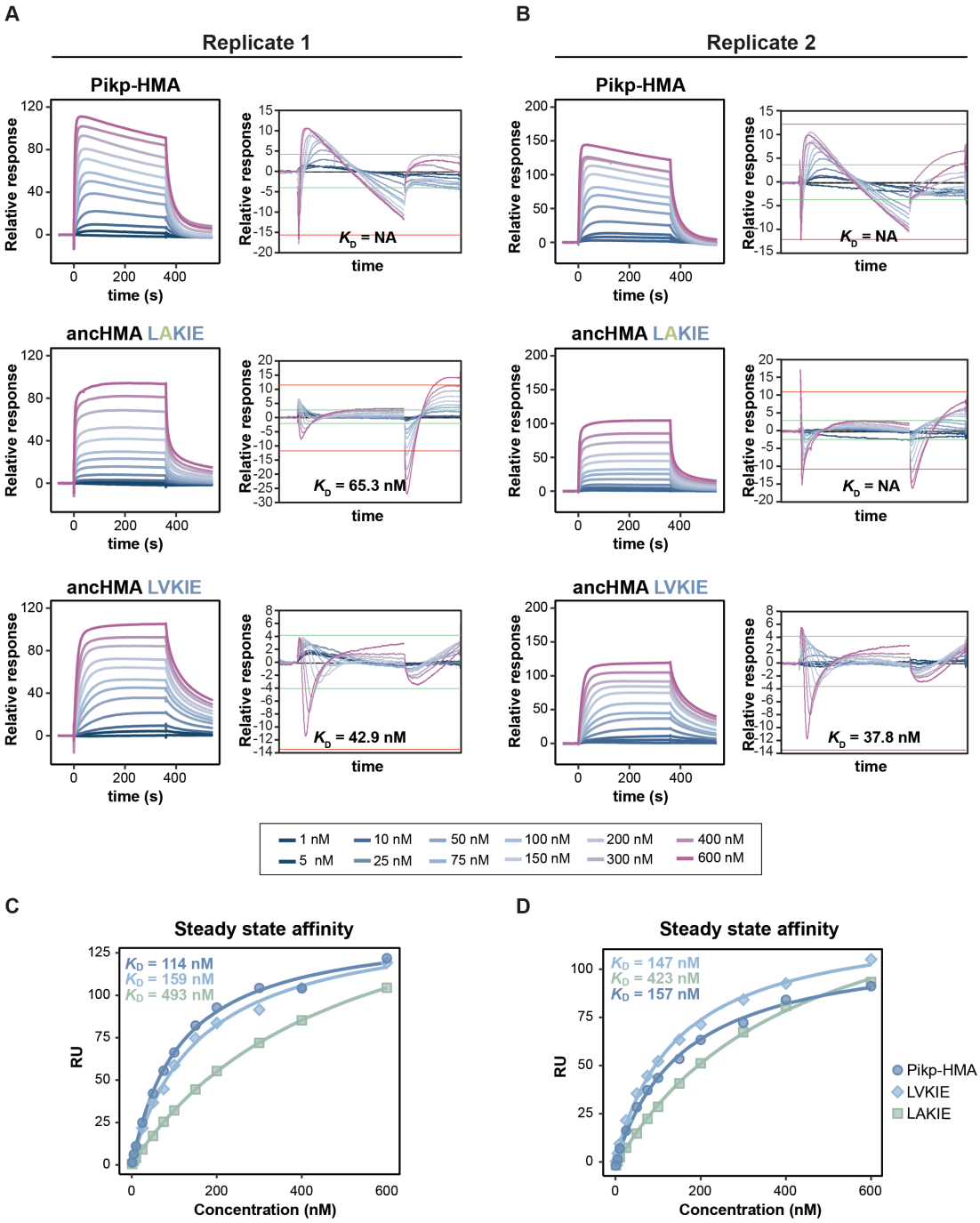


Figure A.3.9 Replicates of multicycle kinetics for the Pikp-HMA, ancHMA$_{LVKIE}$, and ancHMA$_{LAKIE}$ interactions with the AVR-PikD effector, measured by surface plasmon resonance

(**A**,**B**) Multicycle kinetics sensograms illustrating the association and dissociation of AVR-PikD (ligand) and the HMA proteins (analytes) from replicates 1 (**A**) and 2 (**B**). The sensograms for each sample were used to calculate association/dissociation constants (K_D) based on one-to-one kinetics model. Residual plots (right), indicating the difference between the data and the model, were used to access the quality of the fit; SPR acceptance guides as determined by Biacore software

are shown as green and red lines. The K_D values that passed the quality control implemented in Biacore software are shown below respective residual plots; for experiments with insufficient data quality the K_D values are not reported (NA). For all the experiment His-tagged AVR-PikD was immobilised on the sample cell, giving a response level of 157 ± 15 response units (RU). (**C**, **D**) Steady state affinity biding curves, derived from multicycle kinetics sensograms from replicates 1 (**C**) and 2 (**D**), used for K_D calculation.

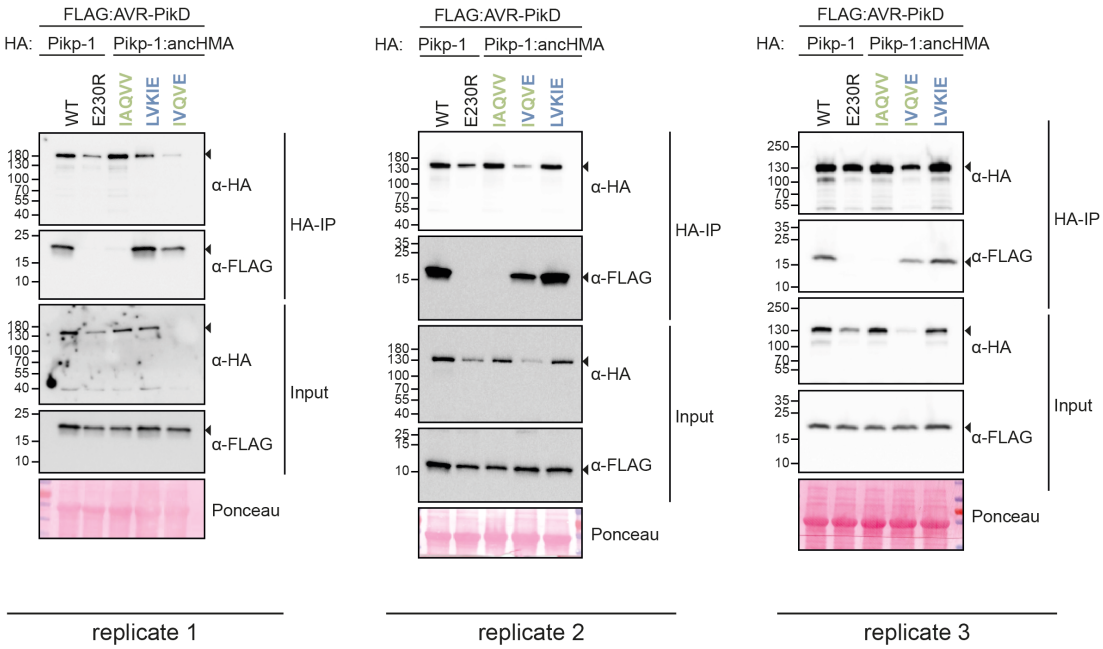


Figure A.3.10 Ala-222-Val and Val-230-Glu substitutions are sufficient to increase binding affinity towards the AVR-PikD effector

Co-immunoprecipitation experiments between FLAG-tagged AVR-PikD with Pikp-1 and Pikp-1:ancHMA constructs featuring HA tag, labelled above. Wild type (WT) HA:Pikp-1 and HA:Pikp-1$_{E230R}$ mutant were used as positive and negative controls, respectively. Proteins, obtained by immunoprecipitation using HA resin (HA-IP), and total protein extracts (Input) were immunoblotted with the appropriate antisera labelled on the right. Loading control, showing rubisco, was performed using Ponceau staining solution. The black arrowheads (right) point to expected band sizes. The figure features results from three independent experiments.

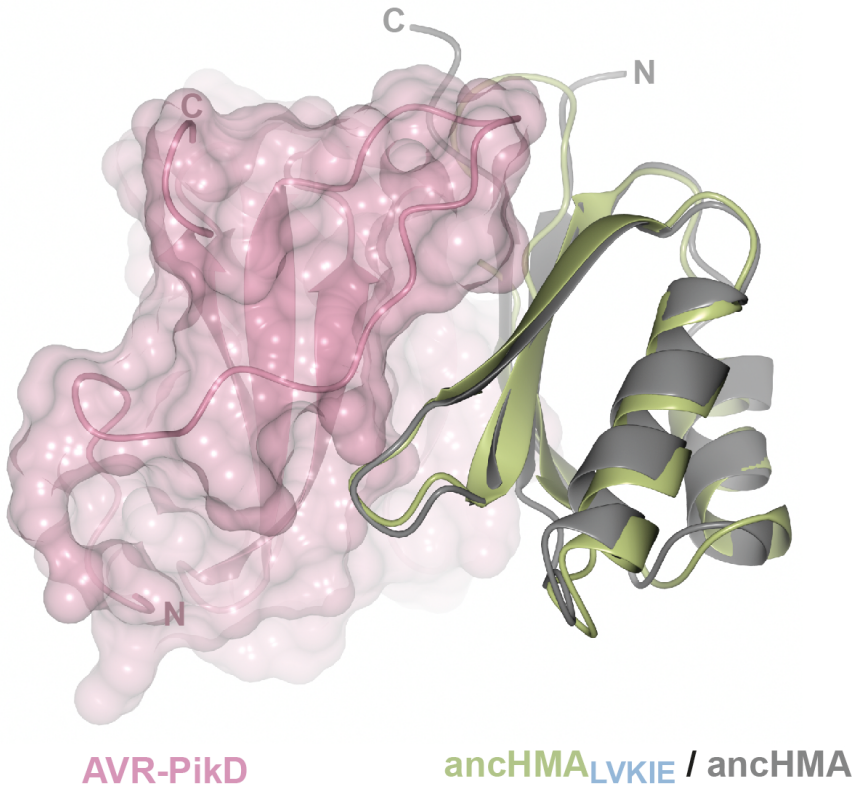
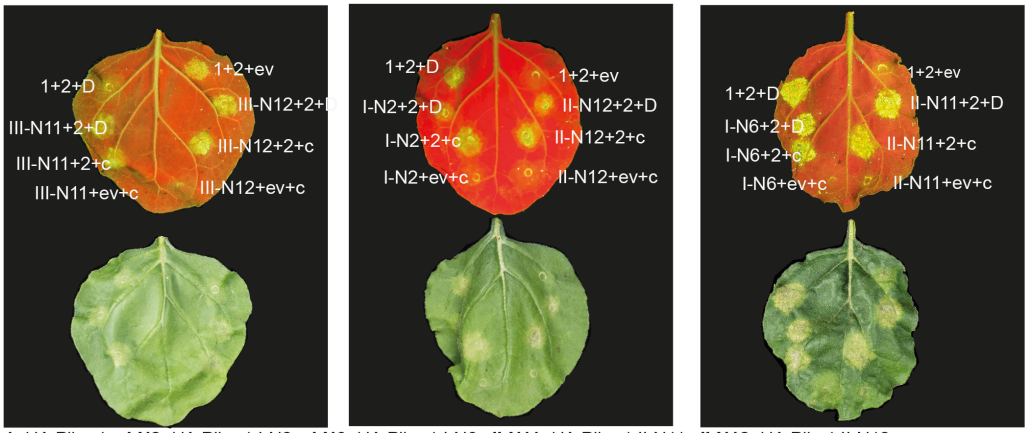


Figure A.3.11 Superimposition of the ancHMA$_{LVKIE}$ structure and the ancHMA homology model in complex with AVR-PikD

Schematic representation of the ancHMA (grey) model superimposed over the ancHMA$_{LVKIE}$ (green) structure with the AVR-PikD (pink) effector. The molecules are shown as ribbons; the molecular surface of AVR-PikD (pink) is also shown. The homology model of the ancHMA–AVR-PikD complex was generated based on the structure of Pikm-HMA in complex with AVR-PikD (pdb: 6fu9).

A



1: HA:Pikp-1; **I-N2:** HA:Pikp-1:I-N2; **I-N6:** HA:Pikp-1:I-N6; **II-N11:** HA:Pikp-1:II-N11; **II-N12:** HA:Pikp1:II-N12;
III-N11: HA:Pikp-1:III-N11; **III-N12:** HA:Pikp-1:III-N12; **2:** Pikp-2:Myc; **D:** AVR-PikD:FLAG; **c:** AVRblb2:FLAG; **ev:** empty vector

B

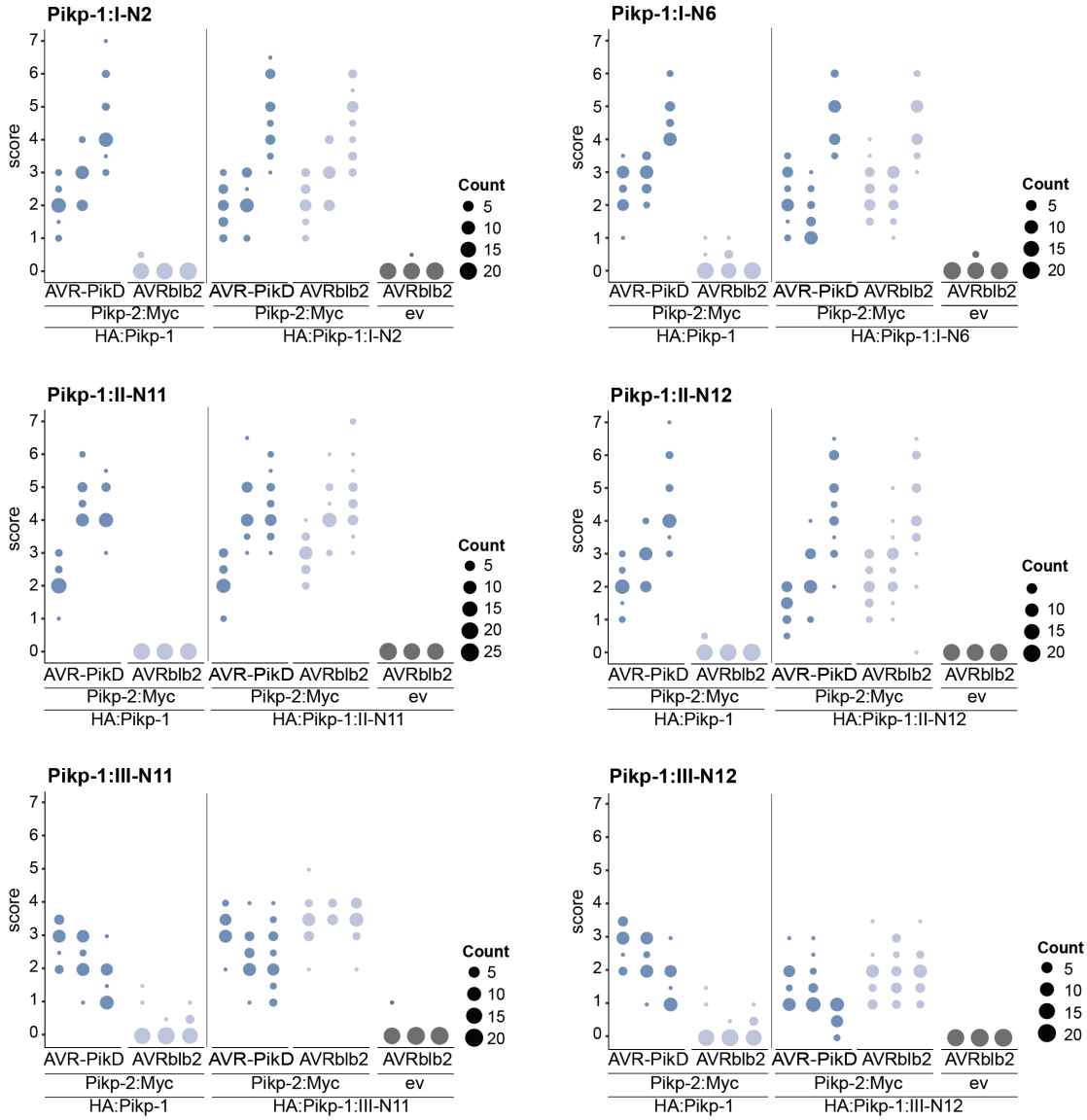


Figure A.3.12 Different Pikp-1:ancHMA fusions are autoactive in Pikp-2–dependent manner

Cell death experiment after transient co-expression of Pikp-1 with the ancestral sequences of HMA with AVR-PikD/AVRblb2 and Pikp-2/empty vector (ev). AVRblb2, an effector from the potato blight pathogen *Phytophthora infestans*, and the empty vector were used as negative controls. Hypersensitive response (HR) was scored five days after agroinfiltration. (**A**) Representative *Nicotiana benthamiana* leaves infiltrated with samples (labelled next to the infiltration spot) were photographed five days post infiltration under UV (top) and day light (bottom). (**B**) HR was scored five days after agroinfiltration. The results are presented as a dots plot, where a size of a dot is proportional to the number of samples with the same score (count) within the same replicate. The experiment was repeated at least three times with 22–26 internal replicates; the columns within tested conditions (labelled on the bottom) show results from different biological replicates.

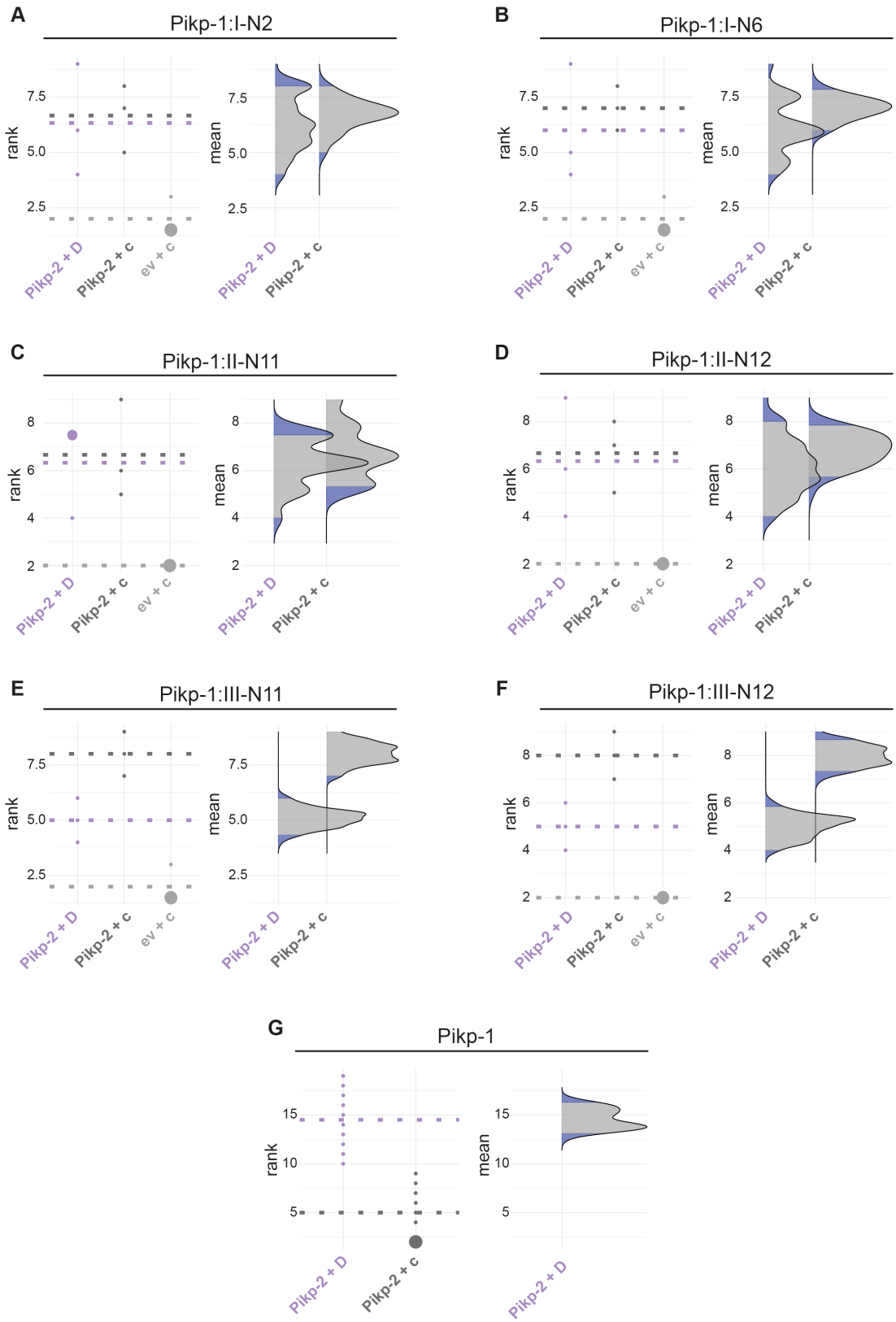
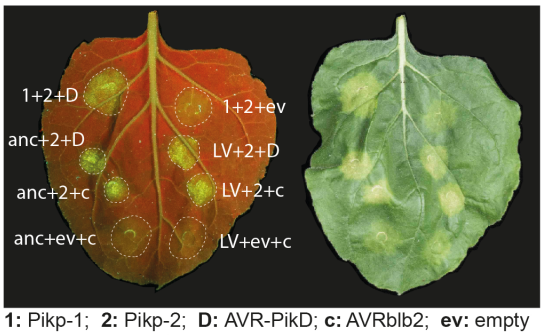

A
Pikp-1:I-N2
rank
mean
7.5
5.0
2.5
Pikp-2 + D
Pikp-2 + c
ev + c
B
Pikp-1:I-N6
C
Pikp-1:II-N11
8
6
4
2
D
Pikp-1:II-N12
E
Pikp-1:III-N11
F
Pikp-1:III-N12
G
Pikp-1
15
10
5

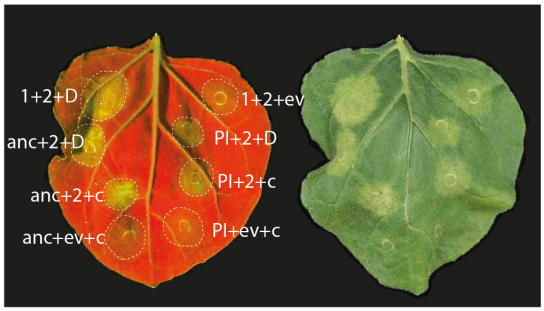
Figure A.3.13 Statistical analysis of cell death for the Pikp-1:ancHMA fusions

The statistical analysis was conducted using an estimation method using besthr R library (MacLean, 2019). (**A–G**) Each panel corresponds to a different HA:Pikp-1:ancHMA fusion (labelled above), co-expressed with Pikp-2:Myc and AVR-PikD:FLAG ('Pikp-2 + D', purple), Pikp-2:Myc and AVRblb2:FLAG ('Pikp-2 + c', dark grey), or empty vector and AVRblb2:FLAG ('ev + c', light grey). AVRblb2:FLAG and empty vector were used as controls. The left panels represent the ranked data (dots) and their corresponding mean (dashed line), with the size of a dot proportional to the number of observations with that specific value. The panels on the right show the distribution of 1000 bootstrap sample rank means, with the blue areas illustrating the 0.025 and 0.975 percentiles of the distribution. The difference is considered significant if the ranked mean for a given condition falls within or beyond the blue percentile of the mean distribution for another condition.

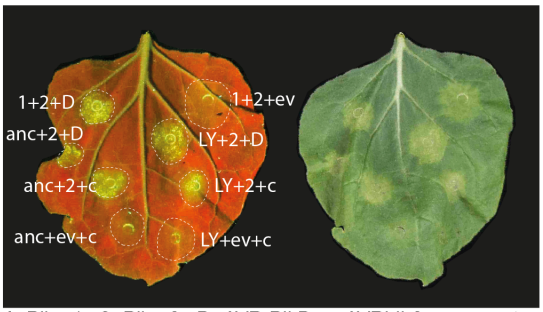
A



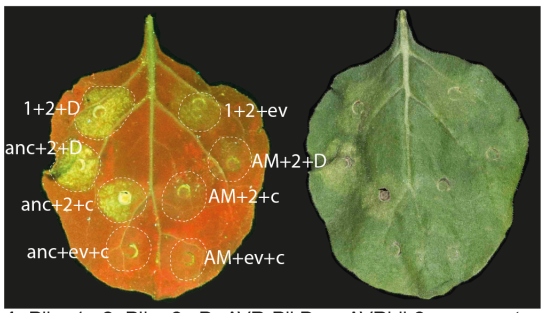
1: Pikp-1; **2:** Pikp-2; **D:** AVR-PikD; **c:** AVRblb2; **ev:** empty vector; **anc:** Pikp-1:ancHMA; **LV:** Pikp-1:ancHMA$_{LVKIE}$



1: Pikp-1; **2:** Pikp-2; **D:** AVR-PikD; **c:** AVRblb2; **ev:** empty vector; **anc:** Pikp-1:ancHMA; **PI:** Pikp-1ancHMA$_{PI}$

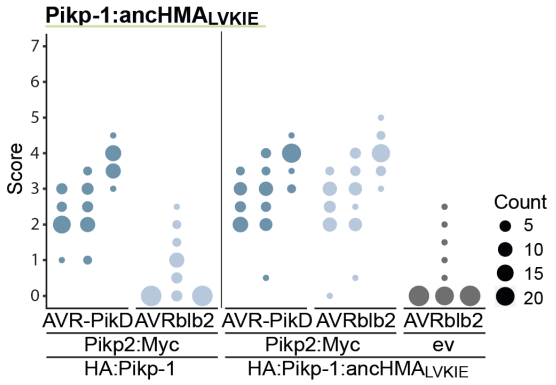


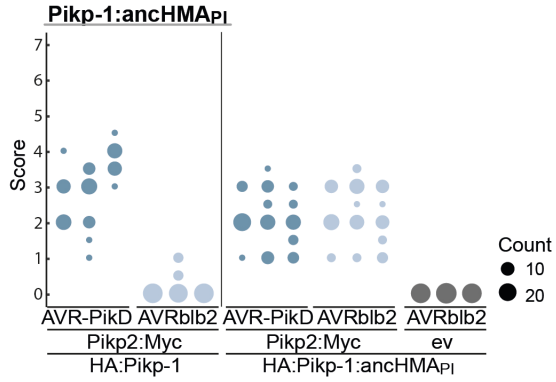
1: Pikp-1; **2:** Pikp-2; **D:** AVR-PikD; **c:** AVRblb2; **ev:** empty vector; **anc:** Pikp-1:ancHMA; **LY:** Pikp-1:ancHMA$_{LY}$

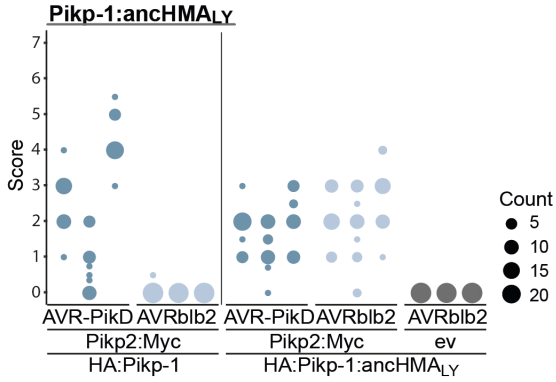


1: Pikp-1; **2:** Pikp-2; **D:** AVR-PikD; **c:** AVRblb2; **ev:** empty vector; **anc:** Pikp-1:ancHMA; **AM:** Pikp-1:ancHMA$_{AMEGNND}$

B







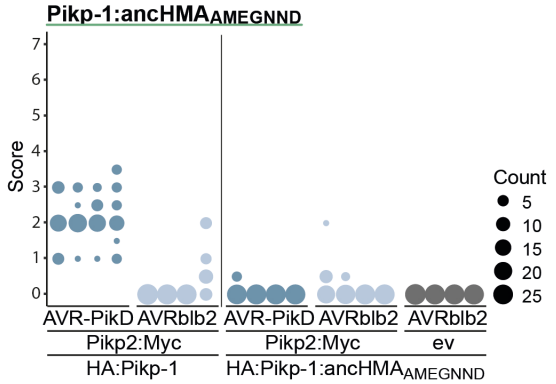


Figure A.3.14 Mutations within the ancHMA abolish autoactivity

Hypersensitive response (HR) assay after transient co-expression of the Pikp-1:ancHMA mutants with AVR-PikD and Pikp-2. AVRblb2, an effector from the potato blight pathogen, and the empty vector (ev) were used as negative controls. (**A**) Representative *N. benthamiana* leaves infiltrated with appropriate constructs (labelled next to the infiltration spot) were photographed five days post-infiltration under UV (left) and day light (right). (**B**) HR was scored five days after agroinfiltration. The results are presented as a dot plots where a size of a dot is proportional to the number of samples with the same score (count) within the same biological replicate. The experiment was independently repeated at least three times with 20–28 internal replicates; the columns within tested conditions (labelled on the bottom) illustrate results from different biological replicates.

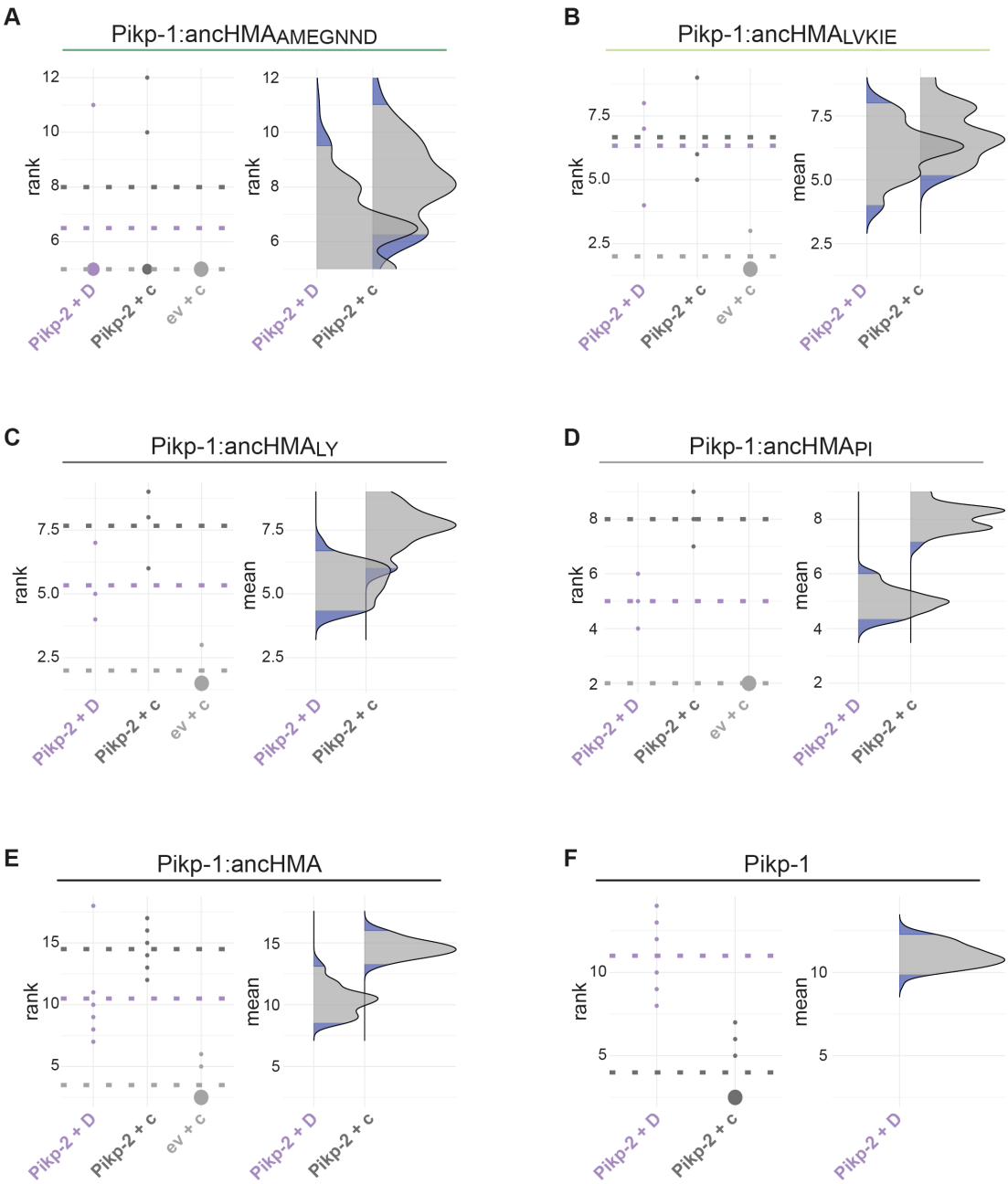


Figure A.3.15 Statistical analysis of cell death assay for the Pikp-1:ancHMA chimeras

The statistical analysis was carried out using an estimation method implemented in besthr R library (MacLean, 2019). (**A–F**) Each panel corresponds to a different chimera of Pikp-1:ancHMA (labelled above), co-expressed with Pikp-2:Myc and AVR-PikD:FLAG ('Pikp-2 + D', purple), Pikp-2:Myc and AVRblb2:FLAG ('Pikp-2 + c', dark grey), or empty vector and AVRblb2:FLAG ('ev + c', light grey). AVRblb2:FLAG and empty vector were used as controls. The left panels represent the ranked data (dots) and their corresponding mean (dashed line), with the size of a dot centre proportional to the number of observations with that specific value. The panels on the right show the distribution of 1000 bootstrap sample rank means, with the blue areas corresponding to the 0.025 and 0.975 percentiles of the distribution. The difference is considered

statistically significant if the ranked mean for a given condition falls within or beyond the blue percentile of the mean distribution for another condition.

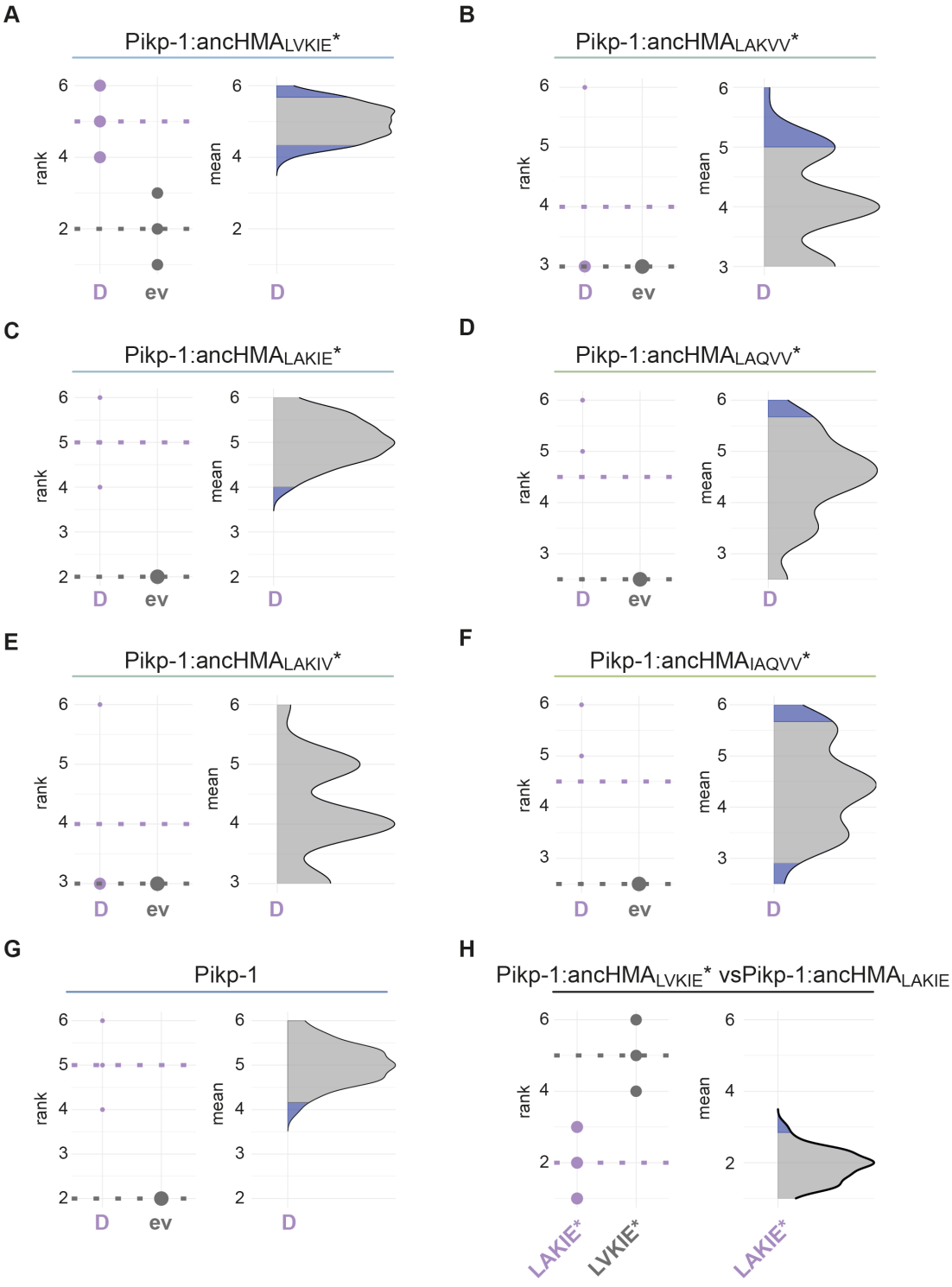


Figure A.3.16 Statistical analysis of cell death for the Pikp-1:ancHMA* mutants

The statistical analysis was performed using an estimation method implemented in besthr R library (MacLean, 2019). (**A–G**) Each panel corresponds to a different Pikp-1:ancHMA mutant (labelled above), featuring HF tag, co-expressed with Myc:AVR-PikD ('D', purple) or empty vector ('ev', grey). All the constructs were co-expressed with Pikp-2. The left panels represent the ranked

data (dots) and their corresponding mean (dashed line). The size of a dot centre is proportional to the number of observations with that specific value. The panels on the right show the distribution of 1000 bootstrap sample rank means, with the blue areas illustrating the 0.025 and 0.975 percentiles of the distribution. The difference is considered significant if the ranked mean for the co-expression with AVR-PikD falls within or beyond the blue percentile of the mean distribution for co-expression with the empty vector. (**H**) Statistical analysis by the estimation method of Pikp:ancHMA$_{LVKIE}$* (LVKIE*) and Pikp:ancHMA$_{LAKIE}$* (LAKIE*) co-expressed with AVR-PikD and Pikp-2 analysed as in panels A–G.

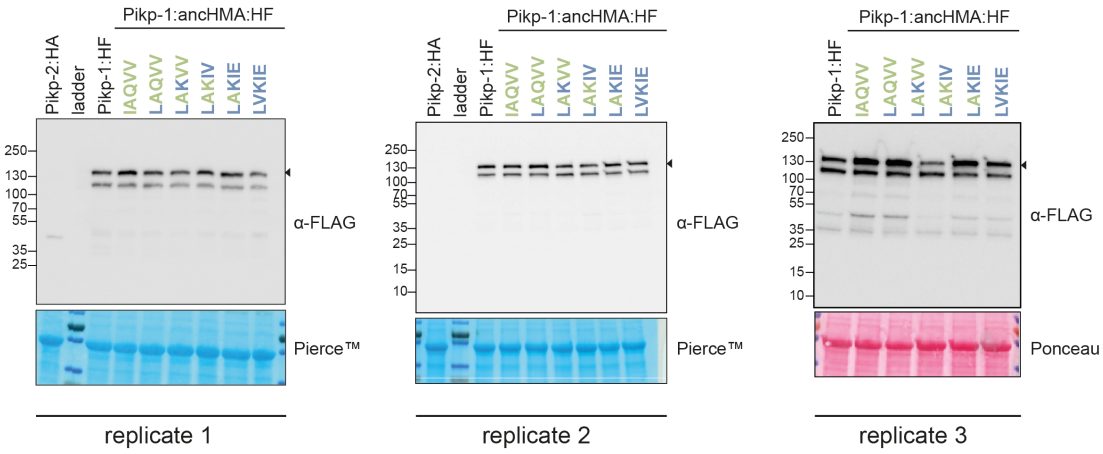


Figure A.3.17 In planta accumulation of the Pikp-1:ancHMA* mutants

Western blot experiments of the Pikp-1:ancHMA* mutants, featuring HF tag, labelled above. Pikp-2:HA construct was included as a negative control. Proteins were immunoblotted with the FLAG antisera (labelled on the right). Rubisco loading controls were performed using Pierce™ or Ponceau staining solutions. The black arrowheads indicate expected band size. The figure features results from three independent experiments.

Table A.3.1 Table of p-values for pairwise comparisons of values for binding and binding stability of in vitro association between AVR-PikD and the HMA mutants in SPR experiments

Sample	Concentration (nM)	Difference	Lower confidence level	Upper confidence level	P-value
E230R	50	-0.2243345	-3.306783	2.858114	0.87441
E230R	200	-0.4872089	-1.217786	0.2433679	1.68E-01
E230R	400	-0.9582807	-1.937141	0.02057976	5.41E-02
IAQVV	50	-1.854604	-2.48629	-1.222918	3.42E-05
IAQVV	200	-6.602886	-7.992786	-5.212985	9.00E-07
IAQVV	400	-12.50061	-15.08403	-9.917196	8.00E-07
LAKIE	50	-6.763288	-9.812373	-3.714204	0.00041
LAKIE	200	-19.31067	-24.76228	-13.85905	1.33E-05
LAKIE	400	-27.75799	-33.91665	-21.59934	1.50E-06
LAKIV	50	-0.7389492	-1.125958	-0.3519399	0.0013223
LAKIV	200	-3.118809	-4.303967	-1.933651	0.0001587
LAKIV	400	-7.148943	-11.29211	-3.005777	3.24E-03
LAKVV	50	-0.3200968	-0.9667166	0.326523	0.3019787
LAKVV	200	-0.4938103	-1.413519	0.4258986	2.59E-01
LAKVV	400	-0.678487	-1.234481	-0.1224934	2.16E-02
LAQVV	50	1.033637	-1.615835	-0.4514394	0.0022349
LAQVV	200	-3.534822	-4.715006	-2.354639	5.54E-05
LAQVV	400	-7.576201	-10.33445	-4.817958	1.13E-04
LVKIE	50	-2.380059	-4.454572	-0.3055451	0.0279298
LVKIE	200	-7.816586	-11.0214	-4.611772	0.0002871
LVKIE	400	-11.5182	-15.92825	-7.10815	0.0001684
Pikp-HMA	50	-7.207938	-8.897001	-5.518875	2.50E-06
Pikp-HMA	200	-17.6612	-20.20803	-15.11436	0
Pikp-HMA	400	-22.40244	-25.83104	-18.97384	0

E230R: Pikp-HMA$_{E230R}$, IAQVV: ancHMA$_{IAQVV}$, LAQVV: ancHMA$_{LAQVV}$, LAKVV: ancHMA$_{LAKVV}$, LAKIV: ancHMA$_{LAKIV}$, LAKIE: ancHMA$_{LAKIE}$, LVKIE: ancHMA$_{LVKIE}$.

Table A.3.2 Table of p-values for all pairwise comparisons of SPR binding to AVR-PikD between the HMA mutants

E230R: Pikp-HMA$_{E230R}$, IAQVV: ancHMA$_{IAQVV}$, LAQVV: ancHMA$_{LAQVV}$, LAKVV: ancHMA$_{LAKVV}$, LAKIV: ancHMA$_{LAKIV}$, LAKIE: ancHMA$_{LAKIE}$, LVKIE: ancHMA$_{LVKIE}$

Concentration (nM)	Sample 1	Sample 2	Difference	Lower confidence level	Upper confidence level	P-value
50	IAQVV	E230R	0.7594754	-6.380544	7.899495	0.9999725
50	LAKIE	E230R	15.1954147	8.055395	22.335434	0.0000005
50	LAKIV	E230R	0.1973327	-6.942687	7.337352	1
50	LAKVV	E230R	-1.2876815	-8.427701	5.852338	0.9990674
50	LAQVV	E230R	-0.2110851	-7.351104	6.928934	1
50	LVKIE	E230R	29.9244059	22.784387	37.064425	0
50	Pikp-HMA	E230R	21.3309677	14.190948	28.470987	0
50	LAKIE	IAQVV	14.4359393	7.29592	21.575959	0.0000016
50	LAKIV	IAQVV	-0.5621428	-7.702162	6.577877	0.9999965
50	LAKVV	IAQVV	-2.047157	-9.187176	5.092862	0.9838216
50	LAQVV	IAQVV	-0.9705606	-8.11058	6.169459	0.9998555
50	LVKIE	IAQVV	29.1649305	22.024911	36.30495	0
50	Pikp-HMA	IAQVV	20.5714923	13.431473	27.711512	0
50	LAKIV	LAKIE	-14.998082	-22.138101	-7.858063	0.0000007
50	LAKVV	LAKIE	-16.483096	-23.623116	-9.343077	0.0000001
50	LAQVV	LAKIE	-15.4065	-22.546519	-8.266481	0.0000004
50	LVKIE	LAKIE	14.7289911	7.588972	21.86901	0.000001
50	Pikp-HMA	LAKIE	6.1355529	-1.004466	13.275572	0.1407186
50	LAKVV	LAKIV	-1.4850142	-8.625033	5.655005	0.9976792
50	LAQVV	LAKIV	-0.4084178	-7.548437	6.731601	0.9999996
50	LVKIE	LAKIV	29.7270732	22.587054	36.867093	0
50	Pikp-HMA	LAKIV	21.133635	13.993616	28.273654	0
50	LAQVV	LAKVV	1.0765964	-6.063423	8.216616	0.9997116
50	LVKIE	LAKVV	31.2120874	24.072068	38.352107	0
50	Pikp-HMA	LAKVV	22.6186492	15.47863	29.758669	0
50	LVKIE	LAQVV	30.135491	22.995472	37.27551	0
50	Pikp-HMA	LAQVV	21.5420528	14.402034	28.682072	0
50	Pikp-HMA	LVKIE	-8.5934382	-15.733457	-1.453419	0.008645
200	IAQVV	E230R	8.12408613	4.769044	11.4791283	0
200	LAKIE	E230R	43.4535131	40.098471	46.8085553	0
200	LAKIV	E230R	7.26935912	3.914317	10.6244013	0.0000006
200	LAKVV	E230R	-0.0357561	-3.390798	3.3192861	1
200	LAQVV	E230R	4.96818917	1.613147	8.3232314	0.0006703
200	LVKIE	E230R	57.0910393	53.735997	60.4460815	0
200	Pikp-HMA	E230R	53.8341313	50.479089	57.1891735	0
200	LAKIE	IAQVV	35.329427	31.974385	38.6844692	0
200	LAKIV	IAQVV	-0.854727	-4.209769	2.5003152	0.9912868
200	LAKVV	IAQVV	-8.1598423	-11.514884	-4.8048001	0
200	LAQVV	IAQVV	-3.155897	-6.510939	0.1991452	0.0782897
200	LVKIE	IAQVV	48.9669532	45.611911	52.3219953	0
200	Pikp-HMA	IAQVV	45.7100452	42.355003	49.0650873	0
200	LAKIV	LAKIE	-36.184154	-39.539196	-32.829112	0

Concentration (nM)	Sample 1	Sample 2	Difference	Lower confidence level	Upper confidence level	P-value
200	LAKVV	LAKIE	-43.489269	-46.844311	-40.134227	0
200	LAQVV	LAKIE	-38.485324	-41.840366	-35.130282	0
200	LVKIE	LAKIE	13.6375262	10.282484	16.9925683	0
200	Pikp-HMA	LAKIE	10.3806182	7.025576	13.7356603	0
200	LAKVV	LAKIV	-7.3051152	-10.660157	-3.9500731	0.0000006
200	LAQVV	LAKIV	-2.30117	-5.656212	1.0538722	0.3776686
200	LVKIE	LAKIV	49.8216802	46.466638	53.1767223	0
200	Pikp-HMA	LAKIV	46.5647722	43.20973	49.9198144	0
200	LAQVV	LAKVV	5.00394529	1.648903	8.3589875	0.0006038
200	LVKIE	LAKVV	57.1267954	53.771753	60.4818376	0
200	Pikp-HMA	LAKVV	53.8698874	50.514845	57.2249296	0
200	LVKIE	LAQVV	52.1228501	48.767808	55.4778923	0
200	Pikp-HMA	LAQVV	48.8659421	45.5109	52.2209843	0
200	Pikp-HMA	LVKIE	-3.256908	-6.61195	0.0981342	0.0625642
400	IAQVV	E230R	15.4160802	10.476265	20.3558957	0
400	LAKIE	E230R	59.1376719	54.197856	64.0774874	0
400	LAKIV	E230R	17.2208188	12.281003	22.1606344	0
400	LAKVV	E230R	-0.6186041	-5.55842	4.3212114	0.9999089
400	LAQVV	E230R	10.5836822	5.643867	15.5234977	0.0000008
400	LVKIE	E230R	67.4451738	62.505358	72.3849893	0
400	Pikp-HMA	E230R	66.0719577	61.132142	71.0117733	0
400	LAKIE	IAQVV	43.7215917	38.781776	48.6614072	0
400	LAKIV	IAQVV	1.8047387	-3.135077	6.7445542	0.9363629
400	LAKVV	IAQVV	-16.034684	-20.9745	-11.094869	0
400	LAQVV	IAQVV	-4.832398	-9.772213	0.1074176	0.0591016
400	LVKIE	IAQVV	52.0290936	47.089278	56.9689091	0
400	Pikp-HMA	IAQVV	50.6558776	45.716062	55.5956931	0
400	LAKIV	LAKIE	-41.916853	-46.856669	-36.977038	0
400	LAKVV	LAKIE	-59.756276	-64.696092	-54.81646	0
400	LAQVV	LAKIE	-48.55399	-53.493805	-43.614174	0
400	LVKIE	LAKIE	8.3075019	3.367686	13.2473174	0.0000904
400	Pikp-HMA	LAKIE	6.9342859	1.99447	11.8741014	0.001416
400	LAKVV	LAKIV	-17.839423	-22.779238	-12.899607	0
400	LAQVV	LAKIV	-6.6371366	-11.576952	-1.6973211	0.0025111
400	LVKIE	LAKIV	50.2243549	45.284539	55.1641705	0
400	Pikp-HMA	LAKIV	48.8511389	43.911323	53.7909544	0
400	LAQVV	LAKVV	11.2022863	6.262471	16.1421019	0.0000002
400	LVKIE	LAKVV	68.0637779	63.123962	73.0035934	0
400	Pikp-HMA	LAKVV	66.6905618	61.750746	71.6303774	0
400	LVKIE	LAQVV	56.8614916	51.921676	61.8013071	0
400	Pikp-HMA	LAQVV	55.4882755	50.54846	60.4280911	0
400	Pikp-HMA	LVKIE	-1.373216	-6.313032	3.5665995	0.9854733

Table A.3.3 Data collection and refinement statistics for the ancHMA$_{LVKIE}$–AVR-PikD co-crystal structure

Property	Value
Data collection statistics	
Wavelength (Å)	0.9700
Space group	P 4_12_12
Cell dimensions	
a, *b*, *c* (Å)	119.49, 119.49, 35.97
α, β, γ, (°)	90.00, 90.00, 90.00
Resolution (Å)*	59.81–1.32 (1.34–1.32)
R_{merge} (%)#	8.4
I/σ(*I*)#	18.7 (1.48)
Completeness (%)#	96.4 (99.8)
Unique reflections#	59,319 (2,981)
Multiplicity#	14.6 (12.3)
$CC^{(1/2)}$ (%)#	99.9 (94.2)
Refinement and model statistics	
Resolution (Å)	59.81-1.32
*R*work/*R*free (%)†	14.7/18.4
No. atoms (Protein)	1784
B-factors (Protein)	22.16
R.m.s. deviations†	
Bond lengths (Å)	0.0129
Bond angles (°)	1.68
Ramachandran plot (%)**	
Favoured	98.64
Outliers	0
MolProbity Score	1.03

*The highest resolution shell is shown in parentheses

**as calculated by MolProbity, #as calculated by Aimless, † as calculated by REFMAC5

Appendix IV

Supplementary information for Chapter 5: AVR-PikD recognition by Pikm-1 evolved through a different path than in Pikp-1—structural basis of convergent molecular evolution

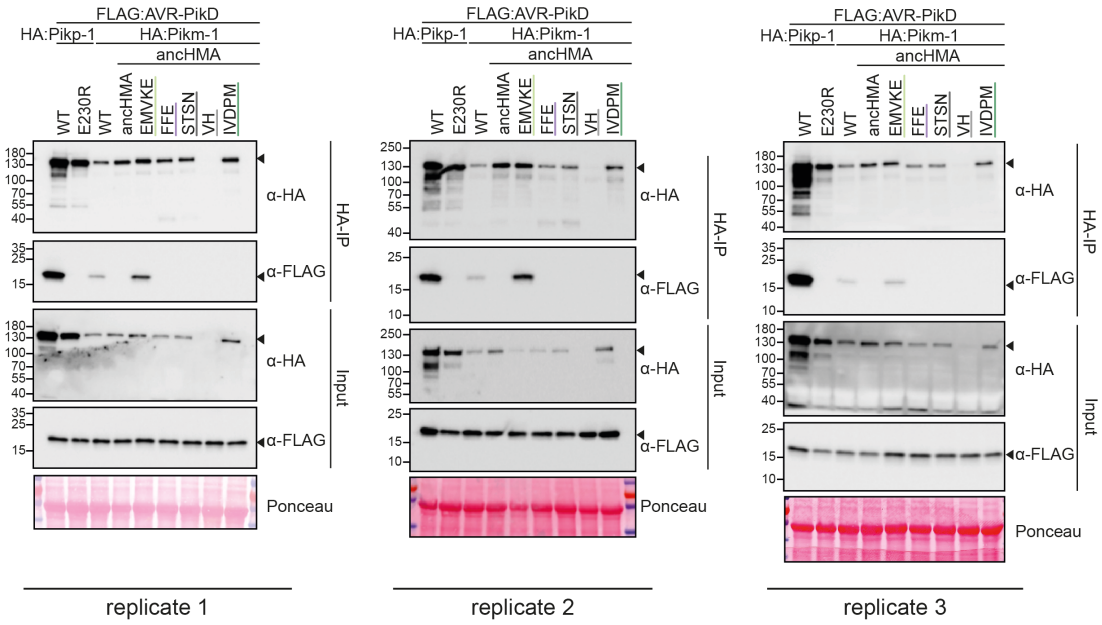


Figure A.4.1 Replicates of the co-immunoprecipitation experiment between the Pikm-1:ancHMA chimeras and AVR-PikD

In planta association of FLAG-tagged AVR-PikD with HA-tagged Pikp-1, Pikp-1_{E230R}, Pikm-1, Pikm-1:ancHMA, and Pikm-1:ancHMA chimeras, labelled above. Wild type (WT) Pikp-1/ Pikm-1 and Pikp-1_{E230R}, featuring HA tag, were used as positive and negative controls, respectively. Proteins obtained by co-immunoprecipitation with HA resin (HA-IP) and total protein extracts (Input) were immunoblotted with the appropriate antisera labelled on the right. Arrowheads indicate expected band sizes. Rubisco loading controls were performed using Ponceau staining solution. The figure presents results from three independent experiments.

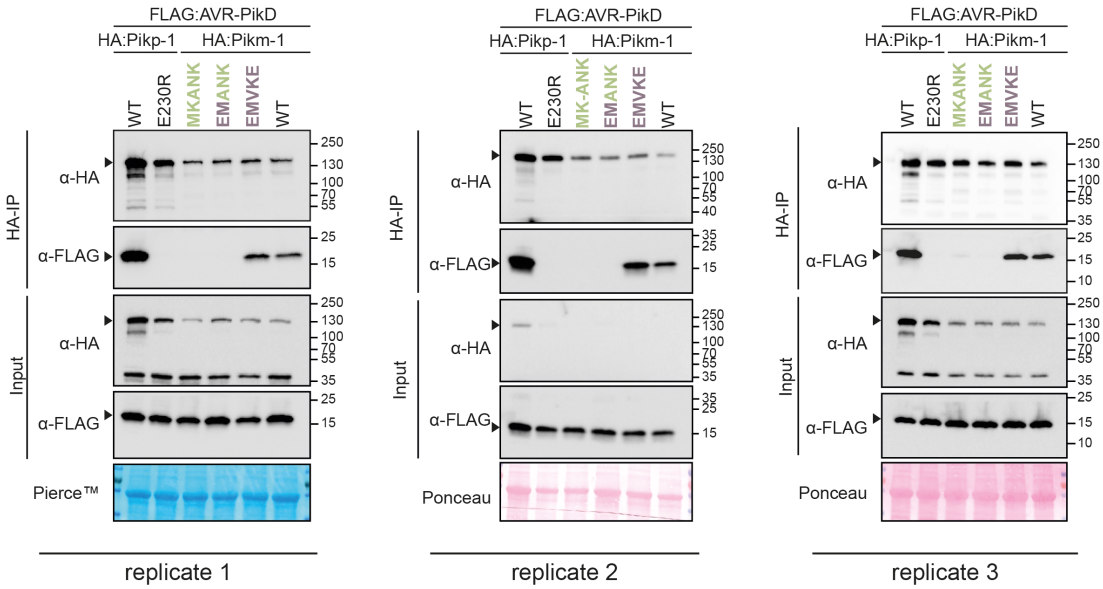
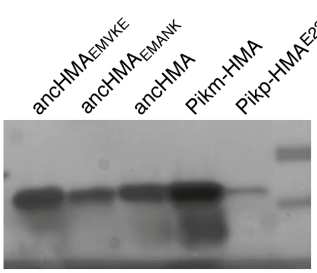


Figure A.4.2 Replicates of the co-immunoprecipitation experiment between the Pikm-1:ancHMA mutants in the EMVKE region and AVR-PikD

In planta association of AVR-PikD, featuring FLAG tag, with HA-tagged Pikp-1, Pikp-1_{E230R}, Pikm-1, Pikm-1:ancHMA, and Pikm-1:ancHMA mutants, labelled above. HA-tagged wild type (WT) Pikp-1/Pikm-1 and Pikp-1_{E230R} were used as positive and negative controls, respectively. Proteins obtained by co-immunoprecipitation with HA-probe (HA-IP) and total protein extracts (Input) were immunoblotted with the appropriate antisera labelled on the left. Arrowheads correspond to expected band sizes. Rubisco loading controls were performed using the Pierce™ or Ponceau staining solutions, labelled on the left. The figure depicts results from three independent experiments.

A



B

Sample name	Predicted mass (Da)	Observed mass (Da)	Difference
ancHMA$_{EMVKE}$	8308.4899	8308.4829	0.8
ancHMA$_{EMANK}$	8265.4590	8265.4484	1.3
ancHMA	8264.5113	8264.4980	0.0133
Pikm-HMA	8531.5420	8531.5440	-0.2
*Pikp-HMA^{E230R}	8445.93 / 8383.6390	8383.6353	62.2947 / 0.4

Figure A.4.3 Purified proteins used in surface plasmon resonance (SPR) studies

(**A**) Coomasie Blue–stained SDS-PAGE gel showing purified HMA proteins used in in vitro experiments. (**B**) Table summarising intact masses of proteins from panel A, as measured by The Sainsbury Laboratory (TSL) core Proteomics facilities (Norwich, UK). (*) The Pikp-HMA^{E230R} protein appears to lack two amino acids at the N-terminus, corresponding to the linker between the protein and the His-tag used for purification. The proteins were purified with help of Adeline Harant from TSL.

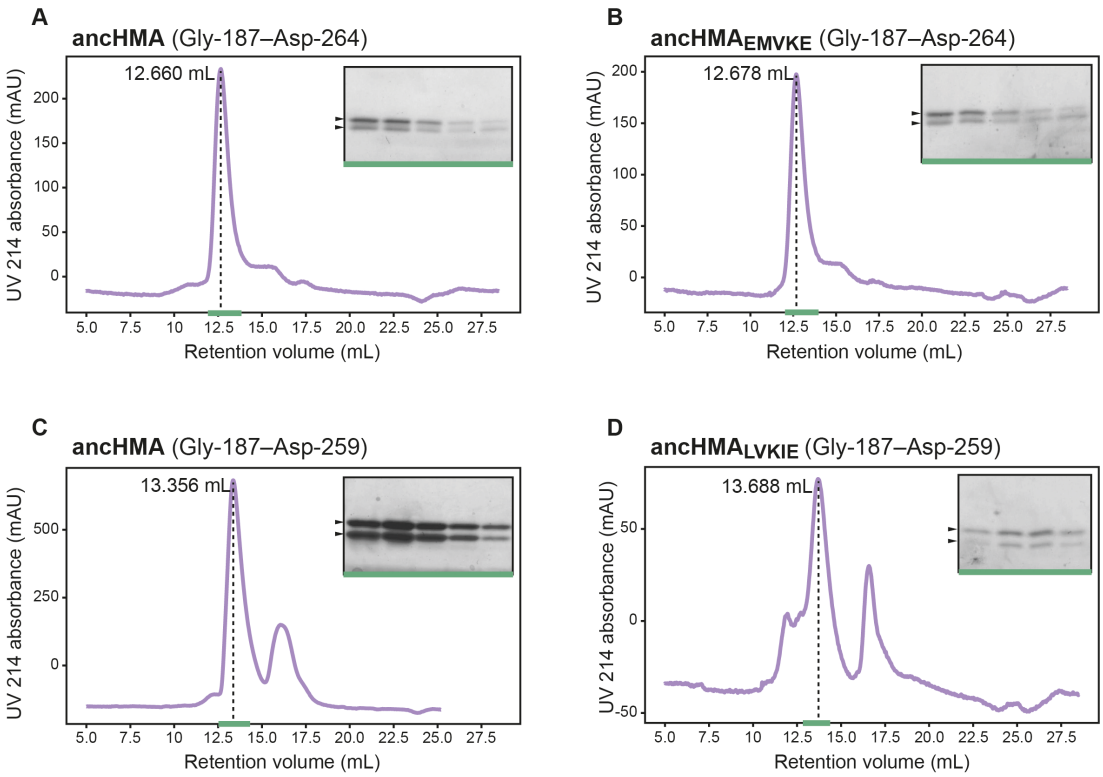


Figure A.4.4 Different stoichiometry of the ancHMA–AVR-PikD complexes

Analytical gel filtration traces depicting the retention volumes of AVR-PikD in complexes with ancHMA (**A**) and ancHMA$_{EMVKE}$ (**B**) with 5–amino acid extension, and ancHMA (**C**) and ancHMA$_{LVKIE}$ (**D**) without the extension. The peakes corresponding to protein complexes are indicated with dahed lines, with the retention volumes shown on the left. Coomasie Blue–stained SDS-PAGE gels of the relevant fractions, marked with green line, are presented of the right. Arrowheads correspond to expected protein sizes.

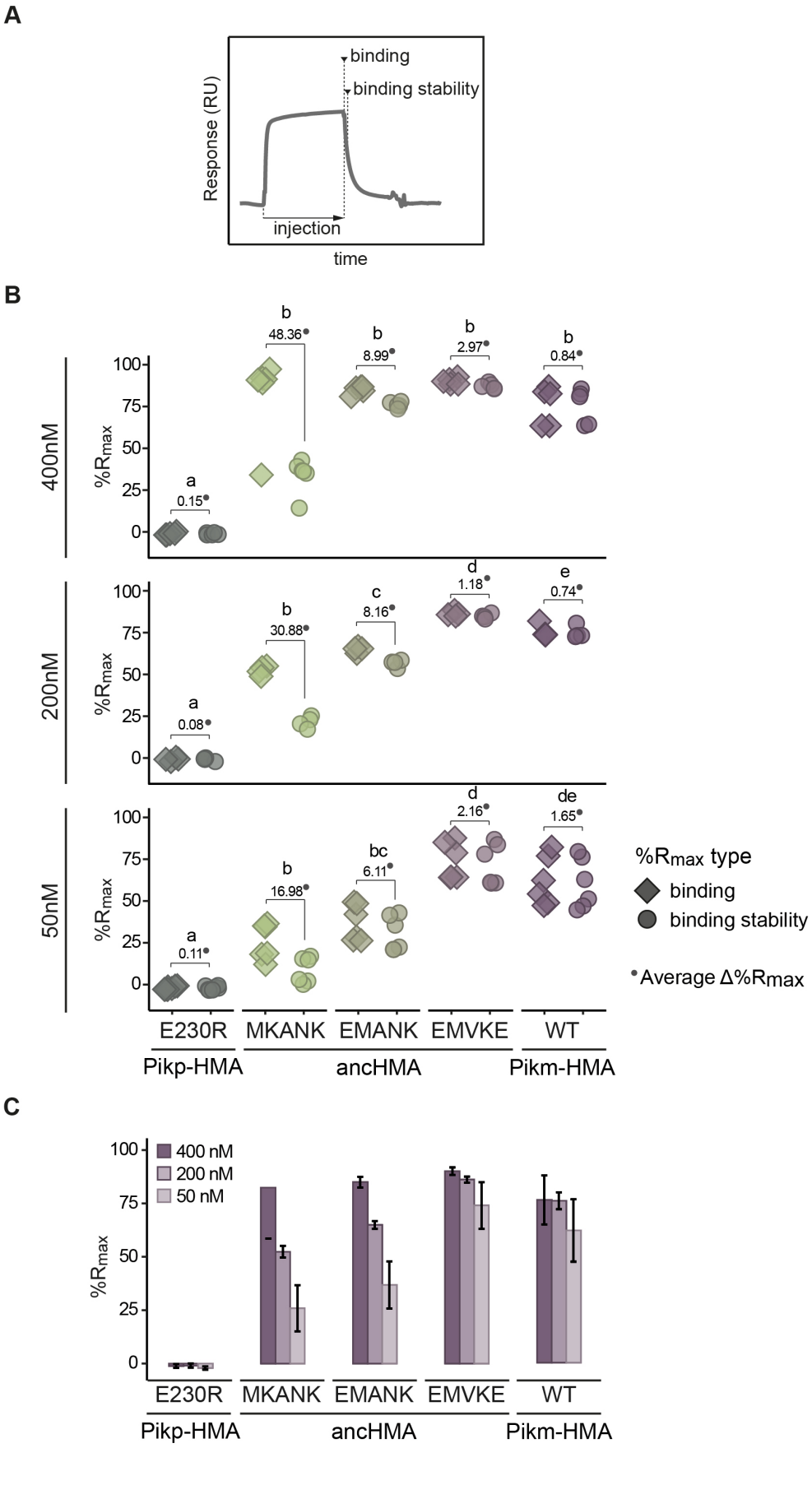

A
binding
binding stability
Response (RU)
injection
time
B
400nM
200nM
50nM
%Rmax
100
75
50
25
0
a
0.15
b
48.36
b
8.99
b
2.97
b
0.84
a
0.08
b
30.88
c
8.16
d
1.18
e
0.74
a
0.11
b
16.98
bc
6.11
d
2.16
de
1.65
E230R
MKANK
EMANK
EMVKE
WT
Pikp-HMA
ancHMA
Pikm-HMA
%Rmax type
binding
binding stability
Average Δ%Rmax
C
400 nM
200 nM
50 nM
%Rmax
100
75
50
25
0
E230R
MKANK
EMANK
EMVKE
WT
Pikp-HMA
ancHMA
Pikm-HMA

Figure A.4.5 Surface plasmon resonance results showing the effect of the EMVKE mutations on AVR-PikD binding in vitro, as indicated by $\%R_{max}$

(**A**) Schematic illustration of the SPR sensogram and the timepoints corresponding to 'binding' and 'binding stability', recorded in this study. (**B**) Plots illustrating calculated percentage of the theoretical maximum response ($\%R_{max}$) values for interaction of the HMA analytes, labelled below, with His-tagged AVR-PikD ligand. $\%R_{max}$ was calculated assuming one-to-one binding model. The values were normalized for the amount of ligand immobilized on the NTA-chip. HMA analytes were tested at three different concentrations, labelled on the left, in at least four independent experiments. All of the data points are represented as diamonds or circles. (•) Average $\Delta\%R_{max}$ values represent absolute differences between average values for 'binding' and 'biding stability' for each sample. Statistical differences among the samples were analysed with ANOVA and Tukey's honest significant difference (HSD) test ($p < 0.01$). P-values for all pairwise comparisons are presented in **Table A.4.2**. (**C**) Mutants of the ancHMA domain display different binding kinetics. The results, identical to those presented in panel B, are shown as histograms to emphasise the differences in binding dynamics between the constructs. Bars represent the average response, and the error bars represent standard deviation.

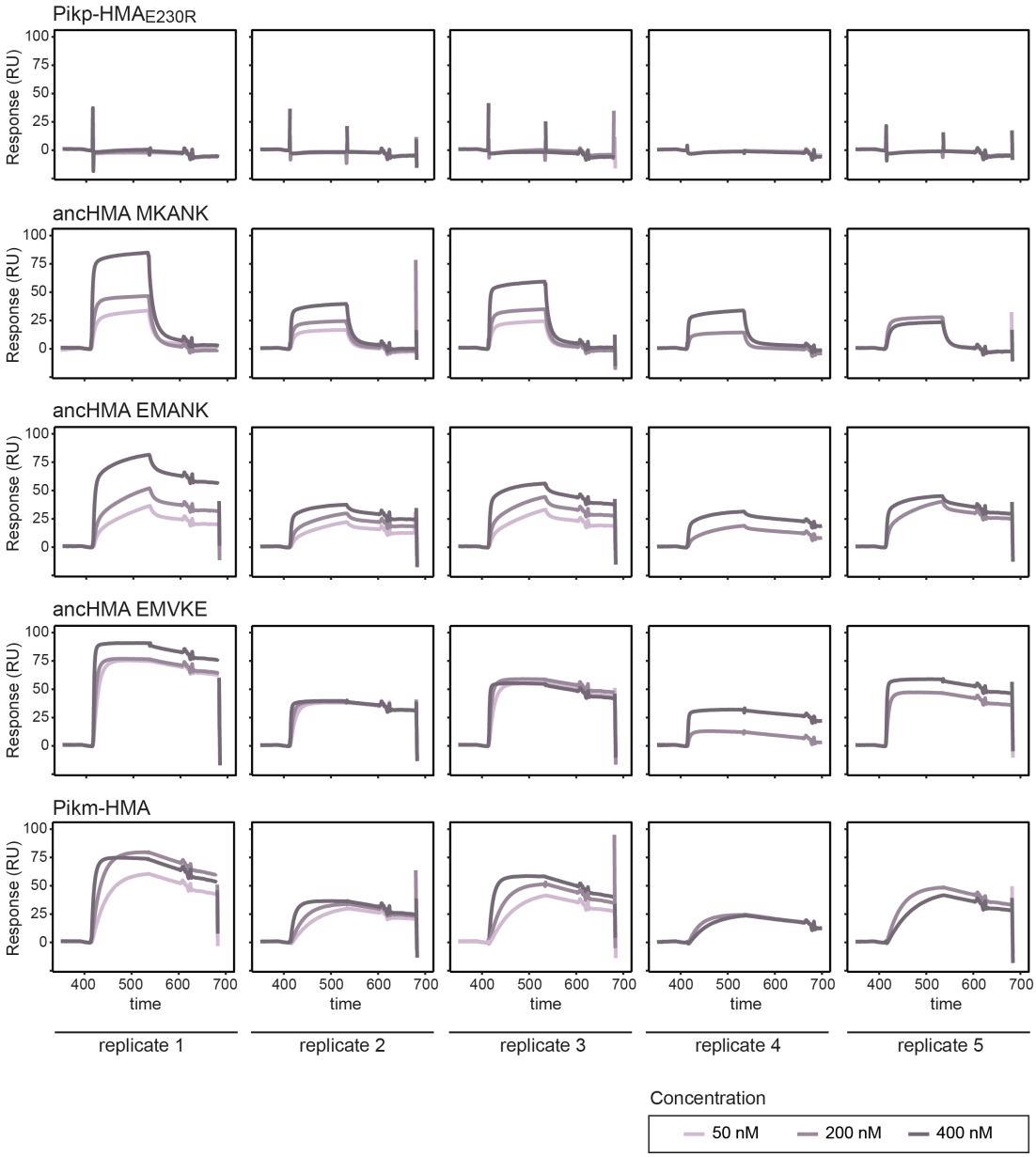


Figure A.4.6 The SPR sensograms for association and dissociation of AVR-PikD and HMA proteins

The SPR sensograms from five independent replicates are shown. His-tagged AVR-PikD was immobilised on the sample cell, giving a response level of 99 ± 33 response units (RU).

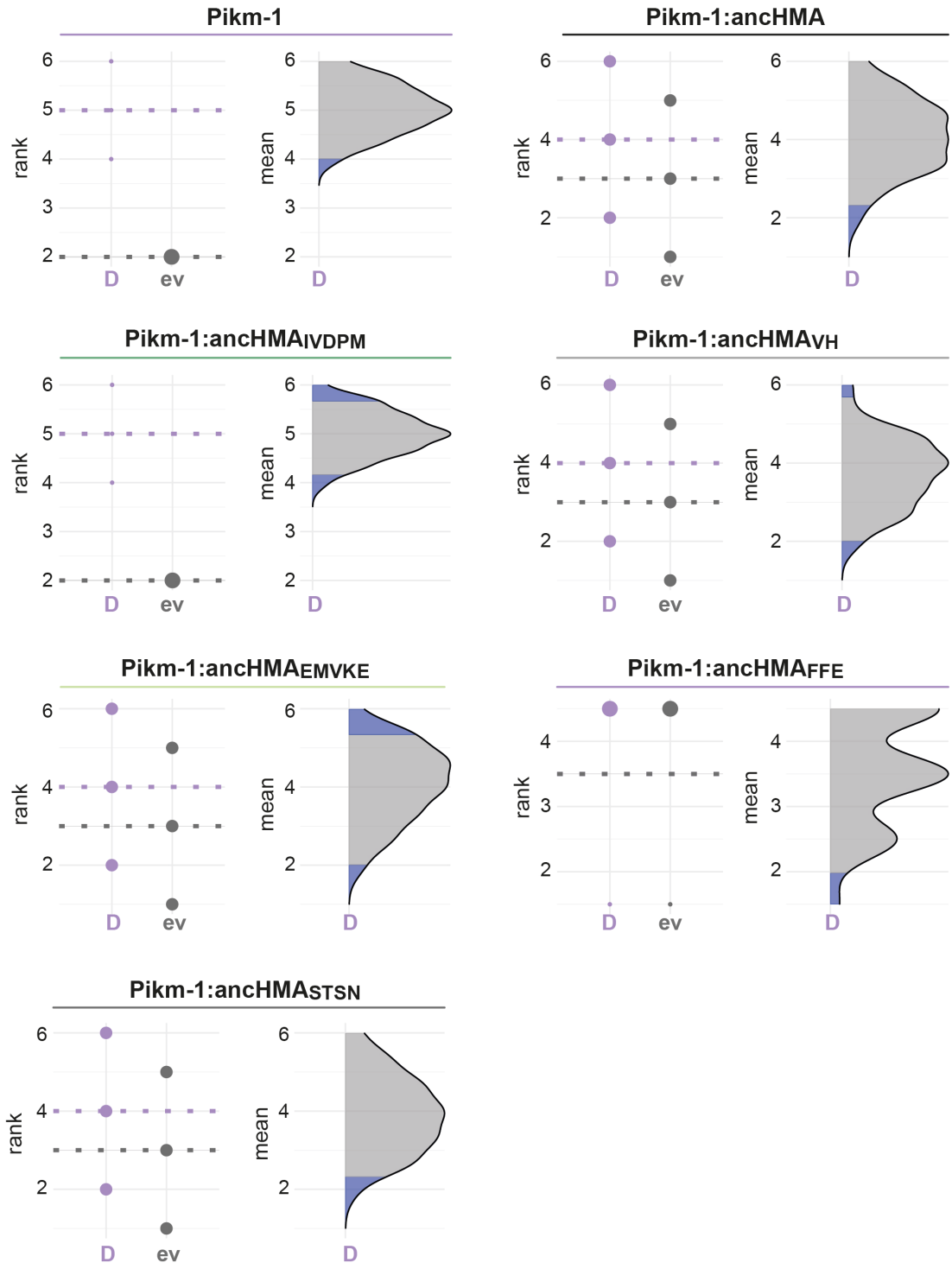


Figure A.4.7 Statistical analysis of cell death for the Pikm-1:ancHMA chimeras

The statistical analysis was performed using estimation methods implemented in the besthr R library (MacLean, 2019). Each panel corresponds to a different HF-tagged Pikm-1:ancHMA mutant (labelled above) co-expressed with Myc:AVR-PikD (D; purple) or empty vector (ev; grey); all constructs were co-expressed with Pikm-2. The left panels represent the ranked data (dots) and their corresponding mean (dashed line). The size of a dot centre is proportional to the number

of observations with that specific value. The panels on the right show the distribution of 1000 bootstrap sample rank means, with the blue areas illustrating the 0.025 and 0.975 percentiles of the distribution. The difference is considered significant if the ranked mean for the co-expression with AVR-PikD falls beyond the blue percentile of the mean distribution for co-expression with the empty vector.

Table A.4.1 Table of p-values for pairwise comparisons of values for binding and binding stability of in vitro association between AVR-PikD and the HMA mutants in SPR experiments

Sample	Concentration (nM)	Difference	Lower confidence level	Upper confidence level	P-value
Pikm-HMA	50	-1.650405	-20.78185	17.48104	0.8514218
EMVKE	50	-0.7392762	-7.403112	5.924559	0.7951379
EMANK	50	-0.8420119	-16.92709	15.24307	0.9068949
MKANK	50	-2.161732	-17.21702	12.89356	0.7556056
E230R	50	-1.182196	-3.693173	1.32878	0.2931335
Pikm-HMA	200	-2.973125	-5.544466	-0.4017852	0.0285241
EMVKE	200	-6.109899	-20.04113	7.82133	0.3515139
EMANK	200	-8.16444	-11.49354	-4.835341	0.0009637
MKANK	200	-8.988376	-12.19905	-5.777706	0.0001971
E230R	200	-16.97771	-29.09641	-4.859003	0.010846
Pikm-HMA	400	-30.877	-36.03862	-25.71538	6.20E-06
EMVKE	400	-48.35823	-71.93835	-24.7781	0.0010271
EMANK	400	-0.1081284	-1.226932	1.010675	0.8367525
MKANK	400	0.07984409	-1.440548	1.600236	0.9019518
E230R	400	-0.073648	-1.146337	0.9990411	0.878125

E230R: Pikp-HMA$_{E230R}$, EMVKE: ancHMA$_{EMVKE}$, EMANK: ancHMA$_{EMANK}$, MKANK: ancHMA

Table A.4.2 Table of p-values for all pairwise comparisons of SPR binding to AVR-PikD between the HMA mutants

Concentration (nM)	Sample 1	Sample 2	Difference	Lower confidence level	Upper confidence level	P-value
400	EMANK	E230R	86.07761	62.02624	110.12898	0
400	EMVKE	E230R	91.209289	67.15792	115.26066	0
400	MKANK	E230R	83.519556	60.49213	106.54699	0
400	Pikm	E230R	77.298931	53.24757	101.3503	0
400	EMVKE	EMANK	5.131679	-18.91969	29.18304	0.967442
400	MKANK	EMANK	-2.558054	-25.58548	20.46938	0.9971982
400	Pikm	EMANK	-8.778679	-32.83004	15.27269	0.8109631
400	MKANK	EMVKE	-7.689733	-30.71716	15.3377	0.8547095
400	Pikm	EMVKE	-13.910358	-37.96172	10.14101	0.4421083
400	Pikm	MKANK	-6.220625	-29.24805	16.8068	0.9262409
200	EMANK	E230R	65.82556	60.553591	71.09752	0
200	EMVKE	E230R	87.06222	81.790259	92.334189	0
200	MKANK	E230R	53.28956	48.017593	58.561522	0
200	Pikm	E230R	76.73675	71.464784	82.008713	0
200	EMVKE	EMANK	21.23667	15.964704	26.508633	0
200	MKANK	EMANK	-12.536	-17.807963	-7.264034	0.0000209
200	Pikm	EMANK	10.91119	5.639228	16.183157	0.0001023
200	MKANK	EMVKE	-33.77267	-39.044631	-28.500702	0
200	Pikm	EMVKE	-10.32548	-15.59744	-5.053511	0.0001865
200	Pikm	MKANK	23.44719	18.175226	28.719156	0
50	EMANK	E230R	38.83454	21.734495	55.934593	0.0000044
50	EMVKE	E230R	76.07891	58.978857	93.178955	0
50	MKANK	E230R	27.8971	10.797048	44.997145	0.0005356
50	Pikm	E230R	63.94274	46.842695	81.042793	0
50	EMVKE	EMANK	37.24436	19.498803	54.989921	0.0000157
50	MKANK	EMANK	-10.93745	-28.683007	6.808111	0.3922477
50	Pikm	EMANK	25.1082	7.362641	42.853759	0.0027261
50	MKANK	EMVKE	-48.18181	-65.927369	-30.436251	0.0000002
50	Pikm	EMVKE	-12.13616	-29.881721	5.609397	0.2926768
50	Pikm	MKANK	36.04565	18.300089	53.791207	0.000026

E230R: Pikp-HMA$_{E230R}$, EMVKE: ancHMA$_{EMVKE}$, EMANK: ancHMA$_{EMANK}$, MKANK: ancHMA; Pikm: Pikm-HMA

References

Adachi H, Contreras MP, Harant A, Wu C, Derevnina L, Sakai T, Duggan C, Moratto E, Bozkurt TO, Maqbool A, Win J, Kamoun S. 2019a. An N-terminal motif in NLR immune receptors is functionally conserved across distantly related plant species. *Elife* 8:e49956.

Adachi H, Derevnina L, Kamoun S. 2019b. NLR singletons, pairs, and networks: evolution, assembly, and regulation of the intracellular immunoreceptor circuitry of plants. *Curr Opin Plant Biol* 50:121–131.

Adachi H, Sakai T, Kourelis J, Maqbool A, Kamoun S. 2020. Jurassic NLR: conserved and dynamic evolutionary features of the atypically ancient immune receptor ZAR1. *bioRxiv* 333484.

Ade J, DeYoung BJ, Golstein C, Innes RW. 2007. Indirect activation of a plant nucleotide binding site-leucine-rich repeat protein by a bacterial protease. *Proc Natl Acad Sci U S A* 104:2531–2536.

Afzal AJ, Kim JH, Mackey D. 2013. The role of NOI-domain containing proteins in plant immune signaling. *BMC Genomics* 14:327.

Alcázar R, García A V., Parker JE, Reymond M. 2009. Incremental steps toward incompatibility revealed by Arabidopsis epistatic interactions modulating salicylic acid pathway activation. *Proc Natl Acad Sci U S A* 106:334–339.

Allen RL, Bittner-Eddy PD, Grenville-Briggs LJ, Meitz JC, Rehmany AP, Rose LE, Beynon JL. 2004. Host-parasite coevolutionary conflict between Arabidopsis and downy mildew. *Science* 306:1957–1960.

Allen RL, Meitz JC, Baumber RE, Hall SA, Lee SC, Rose LE, Beynon JL. 2008. Natural variation reveals key amino acids in a downy mildew effector that alters recognition specificity by an Arabidopsis resistance gene. *Mol Plant Pathol* 9:511–523.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.

Andersen EJ, Nepal MP. 2019. Diversification of disease resistance receptors by integrated domain fusions in wheat and its progenitors. *bioRxiv* 695148.

Aravind L. 2000. The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends Biochem Sci* 25:421–423.

Ashfield T, Ong LE, Nobuta K, Schneider CM, Innes RW. 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, 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.

Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi G, Zomer O, Pupko T. 2012. FastML: A web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res* 40:580–584.

Axtell MJ, Staskawicz BJ. 2003. Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4. *Cell* 112:369–377.

Baggs E, Dagdas G, Krasileva KV. 2017. NLR diversity, helpers and integrated domains: making sense of the NLR IDentity. *Curr Opin Plant Biol* 38:59–67.

Baggs EL, Monroe JG, Thanki AS, O'Grady R, Schudoma C, Haerty W, Krasileva KV. 2020. Convergent loss of an EDS1/PAD4 signaling pathway in several plant lineages reveals coevolved components of plant immunity and drought response. *Plant Cell* 32:2158–2177.

Bai S, Liu J, Chang C, Zhang L, Maekawa T, Wang Q, Xiao W, Liu Y, Chai J, Takken FLW, Schulze-Lefert P, Shen Q-H. 2012. Structure-function analysis of barley NLR immune receptor MLA10 reveals its cell compartment specific activity in cell death and disease resistance. *PLoS Pathog* 8:e1002752.

Bailey PC, Schudoma C, Jackson W, Baggs E, Dagdas G, Haerty W, Moscou M, Krasileva KV. 2018. Dominant integration locus drives continuous diversification of plant immune receptors with exogenous domain fusions. *Genome Biol* 19:23.

Bakker EG, Toomajian C, Kreitman M, Bergelson J. 2006. A genome-wide survey of R gene polymorphisms in Arabidopsis. *Plant Cell* 18:1803–1818.

Banci L, Bertini I, Cantini F, Rosenzweig AC, Yatsunyk LA. 2008. Metal binding domains 3 and 4 of the wilson disease protein: Solution structure and interaction with the copper(I) chaperone HAH1. *Biochemistry* 47:7423–7429.

Barragan AC, Collenberg M, Wang J, Lee RRQ, Cher WY, Rabanal FA, Ashkenazy H, Weigel D, Chae E. 2020. A singleton NLR of recent origin causes hybrid necrosis in *Arabidopsis thaliana*. *bioRxiv* 101451.

Barragan CA, Wu R, Kim S-T, Xi W, Habring A, Hagmann J, Van de Weyer A-L, Zaidem M, Ho WWH, Wang G, Bezrukov I, Weigel D, Chae E. 2019. RPW8/HR repeats control NLR activation in *Arabidopsis thaliana*. *PLOS Genet* 15:e1008313.

Bastolla U, Dehouck Y, Echave J. 2017. What evolution tells us about protein physics, and protein physics tells us about evolution. *Curr Opin Struct Biol* 42:59–66.

Bentham A, Burdett H, Anderson PA, Williams SJ, Kobe B. 2017. Animal NLRs provide structural insights into plant NLR function. *Ann Bot* 119:689–702.

Bergelson J, Kreitman M, Stahl EA, Tian D. 2001. Evolutionary dynamics of plant R-genes. *Science* 292:2281–2285.

Bernoux M, Burdett H, Williams SJ, Zhang X, Chen C, Newell K, Lawrence GJ, Kobe B, Ellis JG, Anderson PA, Dodds PN. 2016. Comparative analysis of the flax immune receptors L6 and L7 suggests an equilibrium-based switch activation Model. *Plant Cell* 28:146–159.

Bernoux M, Ve T, Williams S, Warren C, Hatters D, Valkov E, Zhang X, Ellis JG, Kobe B, Dodds PN. 2011. Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. *Cell Host Microbe* 9:200–211.

Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, Stuart DI, Owens RJ. 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* 35:e45–e45.

Bhullar NK, Zhang Z, Wicker T, Keller B. 2010. Wheat gene bank accessions as a source of new alleles of the powdery mildew resistance gene Pm3: a large scale allele mining project. *BMC Plant Biol* 10:88.

Bialas A, Zess EK, De la Concepcion JC, Franceschetti M, Pennington HG, Yoshida K, Upson JL, Chanclud E, Wu C-H, Langner T, Maqbool A, Varden FA, Derevnina L, Belhaj K, Fujisaki K, Saitoh H, Terauchi R, Banfield MJ, Kamoun S. 2018. Lessons in effector and NLR biology of plant-microbe systems. *Mol Plant-Microbe Interact* 31:34–45.

Bieri S, Mauch S, Shen Q-H, Peart J, Devoto A, Casais C, Ceron F, Schulze S, Steinbiß H-H, Shirasu K, Schulze-Lefert P. 2004. RAR1 Positively Controls Steady State Levels of Barley MLA Resistance Proteins and Enables Sufficient MLA6 Accumulation for Effective Resistance. *Plant Cell* 16:3480–3495.

Blount ZD, Lenski RE, Losos JB. 2018. Contingency and determinism in evolution: Replaying life's tape. *Science* 362:eaam5979.

Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weigel D. 2007. Autoimmune Response as a Mechanism for a Dobzhansky-Muller-Type Incompatibility Syndrome in Plants. PLoS Biol 5:e236.

Bomblies K, Weigel D. 2007. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat Rev Genet* 8:382–393.

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

Bos JIB, Kanneganti T-D, Young C, Cakir C, Huitema E, Win J, Armstrong MR, Birch PRJ, Kamoun S. 2006. The C-terminal half of *Phytophthora infestans* RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in *Nicotiana benthamiana*. *Plant J* 48:165–76.

Boucher JI, Jacobowitz JR, Beckett BC, Classen S, Theobald DL. 2014. An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases. *Elife* 3:1–25.

Boutemy LS, King SRF, Win J, Hughes RK, Clarke TA, Blumenschein TMA, Kamoun S, Banfield MJ. 2011. Structures of *Phytophthora* RXLR effector proteins: A conserved but adaptable fold underpins functional diversity. *J Biol Chem* 286:35834–35842.

Bozkurt TO, Schornack S, Win J, Shindo T, Ilyas M, Oliva R, Cano LM, Jones AME, Huitema E, van der Hoorn R a L, Kamoun S. 2011. *Phytophthora infestans* effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. *Proc Natl Acad Sci U S A* 108:20832–20837.

Brabham HJ, Hernández-Pinzón I, Holden S, Lorang J, Moscou MJ, Moscou1 MJ. 2017. An ancient integration in a plant NLR is maintained as a trans-species polymorphism. *bioRxiv* 23954.

Bridgham JT. 2006. Evolution of hormone-receptor complexity by molecular exploitation. *Science* 312:97–101.

Brown JKM, Tellier A. 2011. Plant-parasite coevolution: Bridging the gap between genetics and ecology. *Annu Rev Phytopathol* 49:345–367.

Brueggeman R, Druka A, Nirmala J, Cavileer T, Drader T, Rostoks N, Mirlohi A, Bennypaul H, Gill U, Kudrna D, Whitelaw C, Kilian A, Han F, Sun Y, Gill K, Steffenson B, Kleinhofs A. 2008. The stem rust resistance gene Rpg5 encodes a protein with nucleotide-binding-site, leucine-rich, and protein kinase domains. *Proc Natl Acad Sci U S A* 105:14970–14975.

Bryan GT, Wu K-S, Farrall L, Jia Y, Hershey HP, McAdams SA, Faulk KN, Donaldson GK, Tarchini R, Valent B. 2000. A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. *Plant Cell* 12:2033–2045.

Bulgarelli D, Biselli C, Collins NC, Consonni G, Stanca AM, Schulze-Lefert P, Valè G. 2010. The CC-NB-LRR-type *Rdg2a* resistance gene confers immunity to the seed-borne barley leaf stripe pathogen in the absence of hypersensitive cell death. *PLoS One* 5:e12599.

Carroll SB. 2005. Evolution at two levels: On genes and form. *PLoS Biol* 3:1159–1166.

Casey LW, Lavrencic P, Bentham AR, Césari S, Ericsson DJ, Croll T, Turk D, Anderson PA, Mark AE, Dodds PN, Mobli M, Kobe B, Williams SJ. 2016. The CC domain structure from the wheat stem rust resistance protein Sr33 challenges paradigms for dimerization in plant NLR proteins. *Proc Natl Acad Sci U S A* 113:12856–12861.

Castel B, Ngou P, Cevik V, Redkar A, Kim D, Yang Y, Ding P, Jones JDG. 2019. Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *New Phytol* 222:966–980.

Césari S. 2018. Multiple strategies for pathogen perception by plant immune receptors. *New Phytol* 219:17–24.

Césari S, Bernoux M, Moncuquet P, Kroj T, Dodds PN. 2014a. A novel conserved mechanism for plant NLR protein pairs: the 'integrated decoy' hypothesis. *Front Plant Sci* 5:606.

Césari S, Kanzaki H, Fujiwara T, Bernoux M, Chalvon V, Kawano Y, Shimamoto K, Dodds P, Terauchi R, Kroj T. 2014b. The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *EMBO J* 33:1941–1959.

Césari S, Thilliez G, Ribot C, Chalvon V, Michel C, Jauneau A, Rivas S, Alaux L, Kanzaki H, Okuyama Y, Morel J-BJ-B, Fournier E, Tharreau D, Terauchi R, Kroj T. 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.

Chae E, Bomblies K, Kim S-T, Karelina D, Zaidem M, Ossowski S, Martín-Pizarro C, Laitinen RAE, Rowan BA, Tenenboim H, Lechner S, Demar M, Habring-Müller A, Lanz C, Rätsch G, Weigel D. 2014. Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. *Cell* 159:1341–1351.

Chae E, Tran DTN, Weigel D. 2016. Cooperation and conflict in the plant immune system. *PLOS Pathog* 12:e1005452.

Chen J, Huang Q, Gao D, Wang J, Lang Y, Liu T, Li B, Bai Z, Luis Goicoechea J, Liang C, Chen C, Zhang W, Sun S, Liao Y, Zhang X, Yang L, Song C, Wang M, Shi J, Liu G, Liu J, Zhou H, Zhou W, Yu Q, An N, Chen Y, Cai Q, Wang B, Liu B, Min J, Huang Y, Wu H, Li Z, Zhang Y, Yin Y, Song W, Jiang J, Jackson S a, Wing R a,

Wang J, Chen M. 2013. Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat Commun* 4:1595.

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr* 66:12–21.

Choi K, Reinhard C, Serra H, Ziolkowski PA, Underwood CJ, Zhao X, Hardcastle TJ, Yelina NE, Griffin C, Jackson M, Mézard C, McVean G, Copenhaver GP, Henderson IR. 2016. Recombination rate heterogeneity within Arabidopsis disease resistance genes. *PLoS Genet* 12:1–30.

Christie N, Tobias PA, Naidoo S, Külheim C. 2016. The Eucalyptus grandis NBS-LRR gene family: Physical clustering and expression hotspots. *Front Plant Sci* 6:1–16.

Christopoulou M, Wo SRC, Kozik A, McHale LK, Truco MJ, Wroblewski T, Michelmore RW. 2015. Genome-wide architecture of disease resistance genes in lettuce. *G3 Genes, Genomes, Genet* 5:2655–2669.

Chung E-H, da Cunha L, Wu A-J, Gao Z, Cherkis K, Afzal AJ, Mackey D, Dangl JL. 2011. Specific threonine phosphorylation of a host target by two unrelated type III effectors activates a host innate immune receptor in plants. *Cell Host Microbe* 9:125–136.

Chung EH, El-Kasmi F, He Y, Loehr A, Dangl JL. 2014. A plant phosphoswitch platform repeatedly targeted by type III effector proteins regulates the output of both tiers of plant immune receptors. *Cell Host Microbe* 16:484–494.

Cohen O, Pupko T. 2011. Inference of gain and loss events from phyletic patterns using stochastic mapping and maximum parsimony-A simulation study. *Genome Biol Evol* 3:1265–1275.

Collier SM, Hamel L-P, Moffett P. 2011. Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Mol Plant-Microbe Interact* 24:918–931.

Collins N, Drake J, Ayliffe M, Sun Q, Ellis J, Hulbert S, Pryor T. 1999. Molecular Characterization of the maize Rp1-D rust resistance haplotype and its mutants. *Plant Cell* 11:1365–1376.

Costanzo S, Jia Y. 2010. Sequence variation at the rice blast resistance gene *Pi-km* locus: Implications for the development of allele specific markers. *Plant Sci* 178:523–530.

Couch BC, Fudal I, Lebrun M-HH, Tharreau D, Valent B, Van Kim P, Nottéghem J-LL, Kohn LM. 2005. Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics* 170:613–630.

De Abreu-Neto JB, Turchetto-Zolet AC, De Oliveira LFV, Bodanese Zanettini MH, Margis-Pinheiro M. 2013. Heavy metal–associated isoprenylated plant protein (HIPP): Characterization of a family of proteins exclusive to plants. *FEBS J* 280:1604–1616.

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

De la Concepcion JC. 2020. Recognition of rice blast effectors by paired immune receptors. John Innes Centre/University of East Anglia.

De la Concepcion JC, Franceschetti M, Maclean D, Terauchi R, Kamoun S, Banfield MJ. 2019. Protein engineering expands the effector recognition profile of a rice NLR immune receptor. *Elife* 8:1–19.

De la Concepcion JC, Franceschetti M, Maqbool A, Saitoh H, Terauchi R, Kamoun S, Banfield MJ. 2018. Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen. *Nat Plants* 4:734–734.

De la Concepcion JC, Maidment JHR, Longya A, Xiao G, Franceschetti M, Banfield MJ. 2020. The allelic rice immune receptor Pikh confers extended resistance to strains of the blast fungus through a single polymorphism in the effector binding interface. *bio*R*xiv* 284240.

De Vries S, Kloesges T, Rose LE. 2015. Evolutionarily dynamic, but robust, targeting of resistance genes by the miR482/2118 gene family in the solanaceae. *Genome Biol Evol* 7:3307–3321.

Dean AM, Thornton JW. 2007. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet* 8:675–688.

Delaux P-M, Hetherington AJ, Coudert Y, Delwiche C, Dunand C, Gould S, Kenrick P, Li F-W, Philippe H, Rensing SA, Rich M, Strullu-Derrien C, de Vries J. 2019. Reconstructing trait evolution in plant evo–devo studies. *Curr Biol* 29:R1110–R1118.

Delaux P-M, Radhakrishnan G V., Jayaraman D, Cheema J, Malbreil M, Volkening JD, Sekimoto H, Nishiyama T, Melkonian M, Pokorny L, Rothfels CJ, Sederoff HW, Stevenson DW, Surek B, Zhang Y, Sussman MR, Dunand C, Morris RJ, Roux C, Wong GK-S, Oldroyd GED, Ané J-M. 2015. Algal ancestor of land plants was preadapted for symbiosis. *Proc Natl Acad Sci U S A* 112:13390–13395.

Delaux P-M, Varala K, Edger PP, Coruzzi GM, Pires JC, Ané JM. 2014. Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLoS Genet* 10: e1004487.

Deng J, Fang L, Zhu X, Zhou B, Zhang T, Zhang D. 2019. A CC-NBS-LRR gene induces hybrid lethality in cotton. *J Exp Bot* 70:5145–5156.

DeSilva TM, Veglia G, Porcelli F, Prantner AM, Opella SJ. 2002. Selectivity in heavy metal-binding to peptides and proteins. *Biopolymers* 64:189–197.

Dodds PN, Lawrence GJ, Catanzariti A-M, Teh T, Wang C-IA, Ayliffe MA, Kobe B, Ellis JG. 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 U S A* 103:8888–8893.

Dodds PN, Lawrence GJ, Catanzariti AM, Ayliffe MA, Ellis JG. 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 PN, Rathjen JP. 2010. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat Rev Genet* 11:539–548.

Dong S, Stam R, Cano LM, Song J, Sklenar J, Yoshida K, Bozkurt TO, Oliva R, Liu Z, Tian M, Win J, Banfield MJ, Jones AME, van der Hoorn RAL, Kamoun S. 2014. Effector specialization in a lineage of the irish potato famine pathogen. *Science* 343:552–555.

Dong Y, Chen S, Cheng S, Zhou W, Ma Q, Chen Z, Fu C-X, Liu X, Zhao Y, Soltis PS, Wong GK-S, Soltis DE, Xiang Q-Y. 2019. Natural selection and repeated patterns of molecular evolution following allopatric divergence. *Elife* 8:e45199.

Duxbury Z, Ma Y, Furzer OJ, Huh SU, Cevik V, Jones JDG, Sarris PF. 2016. Pathogen perception by NLRs in plants and animals: Parallel worlds. *BioEssays* 38:769–781.

Duxbury Z, Wang S, MacKenzie CI, Tenthorey JL, Zhang X, Huh SU, Hu L, Hill L, Ngou PM, Ding P, Chen J, Ma Y, Guo H, Castel B, Moschou PN, Bernoux M, Dodds PN, Vance RE, Jones JDG. 2020. Induced proximity of a TIR signaling domain on a plant-mammalian NLR chimera activates defense in plants. *Proc Natl Acad Sci U S A* 117:18832–18839.

Dyrka W, Coustou V, Daskalov A, Lends A, Bardin T, Berbon M, Kauffmann B, Blancard C, Salin B, Loquet A, Saupe S. 2020. Identification of NLR-associated amyloid signaling motifs in filamentous bacteria. *bioRxiv* 895854.

Dyrka W, Lamacchia M, Durrens P, Kobe B, Daskalov A, Paoletti M, Sherman DJ, Saupe SJ. 2014. Diversity and Variability of NOD-Like Receptors in Fungi. Genome Biol Evol 6:3137–3158.

Eddy SR. 1998. Profile hidden Markov models. *Bioinformatics* 14:755–763.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.

Eitas TK, Dangl JL. 2010. NB-LRR proteins: Pairs, pieces, perception, partners, and pathways. *Curr Opin Plant Biol* 13:472–477.

Ellis JG. 2016. Integrated decoys and effector traps: how to catch a plant pathogen. *BMC Biol* 14:13.

Ellis JG, Lawrence GJ, Luck JE, Dodds PN. 1999. Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. *Plant Cell* 11:495–506.

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr* 66:486–501.

Enkhbayar P, Kamiya M, Osaki M, Matsumoto T, Matsushima N. 2003. Structural principles of leucine-rich repeat (LRR) proteins. *Proteins Struct Funct Bioinforma* 54:394–403.

Essuman K, Summers DW, Sasaki Y, Mao X, Yim AKY, DiAntonio A, Milbrandt J. 2018. TIR Domain Proteins Are an Ancient Family of NAD+-Consuming Enzymes. *Curr Biol* 28:421-430.e4.

Evans PR, Murshudov GN. 2013. How good are my data and what is the resolution? *Acta Crystallogr Sect D Biol Crystallogr* 69:1204–1214.

Fares MA. 2015. The origins of mutational robustness. *Trends Genet* 31:373–381.

Feehan JM, Castel B, Bentham AR, Jones JD. 2020. Plant NLRs get by with a little help from their friends. *Curr Opin Plant Biol* 56:99–108.

Felsenstein J. 1985. Confidence Limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783.

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B. 2003. Map-based isolation of the leaf rust disease resistance gene Lr10 from the hexaploid wheat (*Triticum aestivum* L.) genome. *Proc Natl Acad Sci U S A* 100:15253–15258.

Flor HH. 1971. Current status of the gene-for-gene concept. *Annu Rev Phytopathol* 9:275–296.

Fouché S, Plissonneau C, Croll D. 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Curr Opin Microbiol* 46:34–42.

Franceschetti M, Maqbool A, Jiménez-Dalmaroni MJ, Pennington HG, Kamoun S, Banfield MJ. 2017. Effectors of filamentous plant pathogens: commonalities amid diversity. *Microbiol Mol Biol Rev* 81:1–17.

Fujisaki K, Abe Y, Ito A, Saitoh H, Yoshida K, Kanzaki H, Kanzaki E, Utsushi H, Yamashita T, Kamoun S, Terauchi R. 2015. Rice Exo70 interacts with a fungal effector, AVR-Pii, and is required for AVR-Pii-triggered immunity. *Plant J* 83:875–887.

Fujisaki K, Abe Y, Kanzaki E, Ito K, Utsushi H, Saitoh H, Białas A, Banfield M, Kamoun S, Terauchi R. 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, Ono K, Shimizu T, Ebana K, Hayashi N, Takahashi A, Hirochika H, Okuno K, Yano M. 2009. Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science* 325:998–1001.

Gao Y, Wang W, Zhang T, Gong Z, Zhao H, Han GZ. 2018. Out of water: The ORI in and earl diversification of plant R-genes. *Plant Physiol* 177:82–89.

Germain H, Séguin A. 2011. Innate immunity: has poplar made its BED? New Phytol 189:678–687.

Ginestet C. 2011. ggplot2: Elegant Graphics for Data Analysis. *J R Stat Soc Ser A Statistics Soc* 174:245–246.

Gladieux P, Condon B, Ravel S, Soanes D, Maciel JLN, Nhani A, Chen L, Terauchi R, Lebrun MH, Tharreau D, Mitchell T, Pedley KF, Valent B, Talbot NJ, Farman M, Fournier E. 2018. Gene flow between divergent cereal- and grass-specific lineages of the rice blast fungus *Magnaporthe oryzae*. *MBio* 9:1–19.

Golicz AA, Bayer PE, Barker GC, Edger PP, Kim HR, Martinez PA, Chan CKK, Severn-Ellis A, McCombie WR, Parkin IAP, Paterson AH, Pires JC, Sharpe AG, Tang H, Teakle GR, Town CD, Batley J, Edwards D. 2016. The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat Commun* 7:1–8.

Gómez-Gómez L, 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.

Goritschnig S, Steinbrenner AD, Grunwald DJ, Staskawicz BJ. 2016. Structurally distinct *Arabidopsis thaliana* NLR immune receptors recognize tandem WY domains of an oomycete effector. *New Phytol* 210:984–996.

Grant SR, Fisher EJ, Chang JH, Mole BM, Dangl JL. 2006. Subterfuge and manipulation: Type III effector proteins of phytopathogenic bacteria. *Annu Rev Microbiol* 60:425–449.

Griebel T, Maekawa T, Parker JE. 2014. NOD-like receptor cooperativity in effector-triggered immunity. *Trends Immunol* 35:562–570.

Guo L, Césari S, de Guillen K, Chalvon V, Mammri L, Ma M, Meusnier I, Bonnot F, Padilla A, Peng Y-L, Liu J, Kroj T. 2018. Specific recognition of two MAX effectors by integrated HMA domains in plant immune receptors involves distinct binding surfaces. *Proc Natl Acad Sci U S A* 115:11637–11642.

Gutierrez JR, Balmuth AL, Ntoukakis V, Mucyn TS, Gimenez-Ibanez S, Jones AME, Rathjen JP. 2010. Prf immune complexes of tomato are oligomeric and contain multiple Pto-like kinases that diversify effector recognition. *Plant J* 61:507–518.

Hall S a, Allen RL, Baumber RE, Baxter L a, Fisher K, Bittner-Eddy PD, Rose LE, Holub EB, Beynon JL. 2009. Maintenance of genetic variation in plants and pathogens involves complex networks of gene-for-gene interactions. *Mol Plant Pathol* 10:449–457.

Halterman DA, Wise RP. 2004. A single-amino acid substitution in the sixth leucine-rich repeat of barley MLA6 and MLA13 alleviates dependence on RAR1 for disease resistance signaling. *Plant J* 38:215–226.

Hao W, Collier SM, Moffett P, Chai J. 2013. Structural basis for the interaction between the potato virus X resistance protein (Rx) and its cofactor ran GTPase-activating protein 2 (RanGAP2). *J Biol Chem* 288:35868–35876.

Harms MJ, Thornton JW. 2013. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat Rev Genet* 14:559–571.

Harms MJ, Thornton JW. 2010. Analyzing protein structure and function using ancestral gene reconstruction. *Curr Opin Struct Biol* 20:360–366.

Hartman IV JL. 2001. Principles for the Buffering of Genetic Variation. *Science* 291:1001–1004.

Hayashi K, Yoshida H. 2009. Refunctionalization of the ancient rice blast disease resistance gene Pit by the recruitment of a retrotransposon as a promoter. *Plant J* 57:413–425.

He C, Liang D, Zhang P. 2019. Evaluating the impact of purifying selection on species-level molecular dating. *bioRxiv* 622209.

Heidrich K, Tsuda K, Blanvillain-Baufumé S, Wirthmueller L, Bautor J, Parker JE. 2013. Arabidopsis TNL-WRKY domain receptor RRS1 contributes to temperature-conditioned RPS4 auto-immunity. *Front Plant* Sci 4:403.

Henk AD, Warren RF, Innes RW. 1999. A new Ac-like transposon of Arabidopsis is associated with a deletion of the RPS5 disease resistance gene. *Genetics* 151:1581–1589.

Hodkinson TR. 2018. Evolution and Taxonomy of the Grasses (Poaceae): A Model Family for the Study of Species-Rich Groups, Annual Plant Reviews Online. *Wiley*. 255–294.

Hogenhout SA, Van der Hoorn RAL, Terauchi R, Kamoun S. 2009. Emerging concepts in effector biology of plant-associated organisms. *Mol Plant-Microbe Interact* 22:115–122.

Holub EB. 2001. The arms race is ancient history in Arabidopsis, the wildflower. *Nat Rev Genet* 2:516–527.

Hu Z, Yan C, Liu P, Huang Z, Ma R, Zhang C, Wang R, Zhang Y, Martinon F, Miao D, Deng H, Wang J, Chang J, Chai J. 2013. Crystal structure of NLRC4 reveals its autoinhibition mechanism. *Science* 341:172–175.

Hu Z, Zhou Q, Zhang C, Fan S, Cheng W, Zhao Y, Shao F, Wang H-W, Sui S-F, Chai J. 2015. Structural and biochemical basis for induced self-propagation of NLRC4. *Science* 350:399–404.

Huang J, Si W, Deng Q, Li P, Yang S. 2014. Rapid evolution of avirulence genes in rice blast fungus *Magnaporthe oryzae*. *BMC Genet* 15:15–17.

Huang L, Brooks S, Li W, Fellers J, Nelson JC, Gill B. 2009. Evolution of new disease specificity at a simple resistance locus in a crop-weed complex: Reconstitution of the *Lr21* gene in wheat. *Genetics* 182:595–602.

Hulbert SH, Webb CA, Smith SM, Sun Q. 2001. Resistance gene complexes: evolution and utilization. *Annu Rev Phytopathol* 39:285–312.

Hurni S, Brunner S, Buchmann G, Herren G, Jordan T, Krukowski P, Wicker T, Yahiaoui N, Mago R, Keller B. 2013. Rye Pm8 and wheat Pm3 are orthologous genes and show evolutionary conservation of resistance function against powdery mildew. *Plant J* 76:957–969.

Hurni S, Brunner S, Stirnweis D, Herren G, Peditto D, McIntosh RA, Keller B. 2014. The powdery mildew resistance gene Pm8 derived from rye is suppressed by its wheat ortholog Pm3. *Plant J* 79:904–913.

Inoue Y, Vy TTPP, Yoshida K, Asano H, Mitsuoka C, Asuke S, Anh VL, Cumagun CJRR, Chuma I, Terauchi R, Kato K, Mitchell T, Valent B, Farman M, Tosa Y, Inoue AY, Vy TTPP, Yoshida K, Asano H, Inoue Y, Vy TTPP, Yoshida K, Asano H, Mitsuoka C, Asuke S, Anh VL, Cumagun CJRR, Chuma I, Terauchi R, Kato K, Mitchell T, Valent B, Farman M, Tosa Y. 2017. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. *Science* 357:80–83.

Inukai T, Vales MI, Hori K, Sato K, Hayes PM. 2006. RMo1 confers blast resistance in barley and is located within the complex of resistance genes containing *Mla*, a powdery mildew resistance gene. *Mol Plant-Microbe Interact* 19:1034–1041.

Jacob F, Vernaldi S, Maekawa T. 2013. Evolution and conservation of plant NLR functions. F*ront Immunol* 4:297.

Jia Y, Lee FlN, Mcclung A. 2009. Determination of resistance spectra of the *Pi-ta* and *Pi-k* genes to U.S. races of *Magnaporthe oryzae* causing rice blast in a recombinant inbred line population. *Plant Dis* 93:639–644.

Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B. 2000. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J* 19:4004–4014.

Jiao WB, Schneeberger K. 2020. Chromosome-level assemblies of multiple Arabidopsis genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nat Commun* 11:1–10.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matricies from protein sequences. *Comput Appl Biosci* 8:275–282.

Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* 444:323–329.

Jones JDG, Vance RE, Dangl JL. 2016. Intracellular innate immune surveillance devices in plants and animals. *Science* 354:aaf6395–aaf6395.

Jubic LM, Saile S, Furzer OJ, El Kasmi F, Dangl JL. 2019. Help wanted: helper NLRs and plant immune responses. *Curr Opin Plant Biol* 50:82–94.

Jupe F, Pritchard L, Etherington GJ, MacKenzie K, Cock PJ, Wright F, Sharma SK, Bolser D, Bryan GJ, Jones JD, Hein I. 2012. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics* 13:75.

Kaltenbach M, Burke JR, Dindo M, Pabis A, Munsberg FS, Rabin A, Kamerlin SCL, Noel JP, Tawfik DS. 2018. Evolution of chalcone isomerase from a noncatalytic ancestor article. *Nat Chem Biol* 14:548–555.

Kang S, Sweigard JA, Valent B. 1995. The PWL host specificity gene family in the blast fungus Magnaporthe grisea. *Mol Plant Microbe Interact* 8:939–948.

Kanzaki H, Yoshida K, Saitoh H, Fujisaki K, Hirabuchi A, Alaux L, Fournier E, Tharreau D, Terauchi R. 2012. Arms race co-evolution of *Magnaporthe oryzae* AVR-Pik and rice Pik genes driven by their physical interactions. *Plant J* 72:894–907.

Karasov TL, Kniskern JM, Gao L, Deyoung BJ, Ding J, Dubiella U, Lastra RO, Nallu S, Roux F, Innes RW, Barrett LG, Hudson RR, Bergelson J. 2014. The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature* 512:436–440.

Karasov TL, Shirsekar G, Schwab R, Weigel D. 2020. What natural variation can teach us about resistance durability. *Curr Opin Plant Biol* 56:89–98.

Kessens R, Ashfield T, Kim SH, Innes RW. 2014. Determining the GmRIN4 requirements of the soybean disease resistance proteins Rpg1b and Rpg1r using a *Nicotiana glutinosa*-based agroinfiltration system. *PLoS One* 9:1–7.

Kim S, Park J, Yeom S-I, Kim Y-M, Seo E, Kim K-T, Kim M-S, Lee JM, Cheong K, Shin H-S, Kim S-B, Han K, Lee Jundae, Park M, Lee H-A, Lee Hye-Young, Lee Y, Oh S, Lee JH, Choi Eunhye, Choi Eunbi, Lee SE, Jeon J, Kim H, Choi G, Song H, Lee JunKi, Lee S-C, Kwon J-K, Lee Hea-Young, Koo N, Hong Y, Kim RW, Kang W-H, Huh JH, Kang B-C, Yang T-J, Lee Y-H, Bennetzen JL, Choi D. 2017a. Multiple reference genome sequences of hot pepper reveal the massive evolution of plant

disease resistance genes by retroduplication. *bioRxiv* 115410.

Kim S, Park J, Yeom S, Kim Y-M, Seo E, Kim K-T, Kim M-S, Lee JM, Cheong K, Shin H-S, Kim S-B, Han K, Lee Jundae, Park M, Lee H-A, Lee Hye-Young, Lee Y, Oh S, Lee JH, Choi Eunhye, Choi Eunbi, Lee SE, Jeon J, Kim H, Choi G, Song H, Lee JunKi, Lee S-C, Kwon J-K, Lee Hea-Young, Koo N, Hong Y, Kim RW, Kang W-H, Huh JH, Kang B-C, Yang T-J, Lee Y-H, Bennetzen JL, Choi D. 2017b. New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol* 18:210.

Koenig D, Hagmann J, Li R, Bemm F, Slotte T, Nueffer B, Wright SI, Weigel D. 2019. Long-term balancing selection drives evolution of immunity genes in capsella. *Elife* 8:1–27.

Kourelis J, Kamoun S. 2020. RefPlantNLR: a comprehensive collection of experimentally validated plant NLRs. *bioRxiv* 193961.

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

Krasileva KV, Dahlbeck D, Staskawicz BJ. 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.

Krasileva KV, Zheng C, Leonelli L, Goritschnig S, Dahlbeck D, Staskawicz BJ. 2011. Global analysis of Arabidopsis/downy mildew interactions reveals prevalence of incomplete resistance and rapid evolution of pathogen recognition. *PLoS One* 6:e28765.

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

Krom N, Ramakrishna W. 2008. Comparative analysis of divergent and convergent gene pairs and their expression patterns in rice, Arabidopsis, and *Populus*. *Plant Physiol* 147:1763–1773.

Kuang H, Woo SS, Meyers BC, Nevo E, Michelmore RW. 2004. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell Online* 16:2870–2894.

Kud J, Zhao Z, Du X, Liu Y, Zhao Y, Xiao F. 2013. SGT1 interacts with the Prf resistance protein and is required for Prf accumulation and Prf-mediated defense signaling. *Biochem Biophys Res Commun* 431:501–505.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol* 34:1812–1819.

Kurata N, Yamazaki Y. 2006. Oryzabase. an integrated biological and genome information database for rice. *Plant Physiol* 140:12–17.

Lai Y, Eulgem T. 2018. Transcript-level expression control of plant NLR genes. *Mol Plant Pathol* 19:1267–1281.

Langner T, Harant A, Gomez-Luciano LB, Shrestha RK, Win J, Kamoun S. 2020. Genomic rearrangements generate hypervariable mini-chromosomes in host-specific lineages of the blast fungus. *bioRxiv* 901983.

Latorre SM, Reyes-Avila CS, Malmgren A, Win J, Kamoun S, Burbano HA. 2020. Differential loss of effector genes in three recently expanded pandemic clonal lineages of the rice blast fungus. *BMC Biol* 18:1–15.

Lee H-Y, Mang H, Choi E-H, Seo Y-E, Kim M-S, Oh S, Kim S-B, Choi D. 2019. Genome-wide functional analysis of hot pepper immune receptors reveals an autonomous NLR cluster in seed plants. *bioRxiv* 878959.

Lee S-KS, Song M-Y, Seo Y-S, Kim H-K, Ko S, Cao P-J, Suh J-P, Yi G, Roh J-H, Lee S-KS, An G, Hahn T-R, Wang G-L, Ronald P, Jeon J-S. 2009. Rice Pi5-mediated resistance to *Magnaporthe oryzae* requires the presence of two coiled-coil–nucleotide-binding–leucine-rich repeat genes. *Genetics* 181:1627–1638.

Leister D. 2004. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends Genet* 20:116–122.

Leister TR, Dahlbeck D, Day B, Li Y, Chesnokova O, Staskawicz BJ. 2005. Molecular genetic evidence for the role of SGT1 in the intramolecular complementation of Bs2 protein activity in *Nicotiana benthamiana*. *Plant Cell* 17:1268–1278.

Lenski RE, Barrick JE, Ofria C. 2006. Balancing robustness and evolvability. *PLoS Biol* 4:e428.

Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127–128.

Lewis JD, Lee AHY, Hassan JA, Wana J, Hurleya B, Jhingree JR, Wang PW, Lo T, Youn JY, Guttman DS, Desveaux D. 2013. The Arabidopsis ZED1 pseudokinase is required for ZAR1-mediated immunity induced by the *Pseudomonas syringae* type III effector HopZ1a. *Proc Natl Acad Sci U S A* 110:18722–18727.

Li J, Ding J, Zhang W, Zhang Y, Tang P, Chen JQ, Tian D, Yang S. 2010. Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. *Mol Genet Genomics* 283:427–438.

Li J, Wang Q, Li C, Bi Y, Fu X, Wang R. 2019. Novel haplotypes and networks of AVR-Pik alleles in *Magnaporthe oryzae*. *BMC Plant Biol* 19:1–12.

Li J, Zhang M, Sun J, Mao X, Wang J, Liu H, Zheng H, Li X, Zhao H, Zou D. 2020. Heavy metal stress-associated proteins in rice and Arabidopsis: genome-wide identification, phylogenetics, duplication, and expression profiles analysis. *Front Genet* 11:1–21.

Li L, Habring A, Wang K, Weigel D. 2020. Atypical resistance protein RPW8/HR triggers oligomerization of the NLR immune receptor RPP7 and autoimmunity. *Cell Host Microbe* 27:405-417.e6.

Liao J, Huang H, Meusnier I, Adreit H, Ducasse A, Bonnot F, Pan L, He X, Kroj T, Fournier E, Tharreau D, Gladieux P, Morel J-B. 2016. Pathogen effectors and plant immunity determine specialization of the blast fungus to rice subspecies. *Elife* 5:1–18.

Lin X, Zhang Y, Kuang H, Chen J. 2013. Frequent loss of lineages and deficient duplications accounted for low copy number of disease resistance genes in Cucurbitaceae. *BMC Genomics* 14:335.

Lindbo JA. 2007. TRBO: A high-efficiency tobacco mosaic virus RNA-based overexpression vector. *Plant Physiol* 145:1232–1240.

Liu W, Frick M, Huel R, Nykiforuk CL, Wang X, Gaudet DA, Eudes F, Conner RL, Kuzyk A, Chen Q, Kang Z, Laroche A. 2014. the stripe rust resistance gene *Yr10* encodes an evolutionary-conserved and unique CC–NBS–LRR sequence in wheat. *Mol Plant* 7:1740–1755.

Liu Z, Bos JIB, Armstrong M, Whisson SC, da Cunha L, Torto-Alalibo T, Win J, Avrova AO, Wright F, Birch PRJ, Kamoun S. 2005. Patterns of diversifying selection in the phytotoxin-like scr74 gene family of *Phytophthora infestans*. *Mol Biol Evol* 22:659–672.

Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M. 2012. SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb Cell Fact* 11:753.

Longya A, Chaipanya C, Franceschetti M, Maidment JHR, Banfield MJ, Jantasuriyarat C. 2019. Gene duplication and mutation in the emergence of a novel aggressive allele of the AVR-Pik effector in the rice blast fungus. *Mol Plant-Microbe Interact* 32:740–749.

Losos JB. 2011. Convergence, adaptation, and constraint. *Evolution* 65:1827–40.

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

Lukasik-Shreepaathy E, Slootweg E, Richter H, Goverse A, Cornelissen BJC, Takken FLW. 2012. Dual regulatory roles of the extended n terminus for activation of the tomato Mi-1.2 resistance protein. *Mol Plant-Microbe Interact* 25:1045–1057.

Luo S, Zhang Y, Hu Q, Chen J, Li K, Lu C, Liu H, Wang W, Kuang H. 2012. Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family. *Plant Physiol* 159:197–210.

Ma Y, Guo H, Hu L, Martinez PP, Moschou PN, Cevik V, Ding P, Duxbury Z, Sarris PF, Jones JDG. 2018. Distinct modes of derepression of an Arabidopsis immune receptor complex by two different bacterial effectors. *Proc Natl Acad Sci U S A* 115:10218–10227.

Macho AP, Zipfel C. 2015. Targeting of plant pattern recognition receptor-triggered immunity by bacterial type-III secretion system effectors. *Curr Opin Microbiol* 23:14–22.

MacLean D. 2019. Besthr. *Zenodo*.

MacQueen A, Sun X, Bergelson J. 2016. Genetic architecture and pleiotropy shape costs of Rps2-mediated resistance in Arabidopsis thaliana. Nat Plants 2:16110.

Maekawa T, Cheng W, Spiridon LN, Töller A, Lukasik E, Saijo Y, Liu P, Shen Q-H, Micluta M a, Somssich IE, Takken FLW, Petrescu A-J, Chai J, Schulze-Lefert P. 2011. Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. *Cell Host Microbe* 9:187–99.

Maekawa T, Kracher B, Saur IML, Yoshikawa-Maekawa M, Kellner R, Pankin A, von Korff M, Schulze-Lefert P. 2019. Subfamily-specific specialization of RGH1/MLA immune receptors in wild barley. *Mol Plant-Microbe Interact* 32:107–119.

Mago R, Zhang P, Vautrin S, Šimková H, Bansal U, Luo M-C, Rouse M, Karaoglu H, Periyannan S, Kolmer J, Jin Y, Ayliffe MA, Bariana H, Park RF, McIntosh R, Doležel J, Bergès H, Spielmeyer W, Lagudah ES, Ellis JG, Dodds PN. 2015. The wheat *Sr50* gene reveals rich diversity at a cereal disease resistance locus. *Nat Plants* 1:15186.

Maidment JHR. 2020. Using a rice blast effector target to engineer NLR immune receptors with novel recognition specificities [Unpublished PhD thesis]. John Innes Centre/University of East Anglia.

Maqbool A, Saitoh H, Franceschetti M, Stevenson C, Uemura A, Kanzaki H, Kamoun S, Terauchi R, Banfield M. 2015. Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *Elife* 4:1–24.

Marchal C, Zhang J, Zhang P, Fenwick P, Steuernagel B, Adamski NM, Boyd L, McIntosh R, Wulff BBH, Berry S, Lagudah E, Uauy C. 2018. BED-domain-containing immune receptors confer diverse resistance spectra to yellow rust. *Nat Plants* 4:662–668.

Martin R, Qi T, Zhang H, Liu F, King M, Toth C, Staskawicz BJ. 2020. Structure of the activated Roq1 resistosome directly recognizing the pathogen effector XopQ. *bio*R*xiv* 246413.

McCann HC. 2020. Skirmish or war: the emergence of agricultural plant pathogens. *Curr Opin Plant Biol* 56:147–152.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* 40:658–674.

McDowell JM, Dhandaydham M, Long TA, Aarts MGM, Goff S, Holub EB, Dangl JL. 1998. Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsis. *Plant Cell* 10:1861–1874.

McHale L, Tan X, Koehl P, Michelmore RW. 2006. Plant NBS-LRR proteins: Adaptable guards. *Genome Biol* 7:212.

Mestre P, Baulcombe DC. 2006. Elicitor-mediated oligomerization of the tobacco N disease resistance protein. *Plant Cell* 18:491–501.

Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW. 1998. The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell* 10:1817–1832.

Meyers BC, Kaushik S, Nandety RS. 2005. Evolving disease resistance genes. *Curr Opin Plant Biol* 8:129–134.

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. 2003. Genome-wide analysis of NBS-LRR – encoding genes in Arabidopsis. *Plant Cell* 15:809–834.

Michelmore RW, Meyers BC. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res* 8:1113–1130.

Mizuno H, Katagiri S, Kanamori H, Mukai Y, Sasaki T, Matsumoto T, Wu J. 2020. Evolutionary dynamics and impacts of chromosome regions carrying R-gene clusters in rice. *Sci Rep* 10:1–2.

Moffett P, Farnham G, Peart J, Baulcombe DC. 2002. Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death. *EMBO J* 21:4511–4519.

Moscou M. 2019. QKphylogeny. github.com.//github.com/matthewmoscou/QKphylogeny.

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr Sect D Biol Crystallogr* 67:355–367.

Nandety RS, Caplan JL, Cavanaugh K, Perroud B, Wroblewski T, Michelmore RW, Meyers BC. 2013. The role of TIR-NBS and TIR-X proteins in plant basal defense responses. *Plant Physiol* 162:1459–1472.

Narusaka M, Kubo Y, Hatakeyama K, Imamura J, Ezura H, Nanasato Y, Tabei Y, Takano Y, Shirasu K, Narusaka Y. 2013. interfamily transfer of dual NB-LRR genes confers resistance to multiple pathogens. *PLoS One* 8:e55954.

Narusaka M, Shirasu K, Noutoshi Y, Kubo Y, Shiraishi T, Iwabuchi M, Narusaka Y. 2009. RRS1 and RPS4 provide a dual Resistance-gene system against fungal and bacterial pathogens. *Plant J* 60:218–226.

Nguyen V, Wilson C, Hoemberger M, Stiller JB, Agafonov R V, Kutter S, English J, Theobald DL, Kern D. 2017. Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science* 355:289–294.

Nishimura MT, Anderson RG, Cherkis KA, Law TF, Liu QL, Machius M, Nimchuk ZL, Yang L, Chung EH, El Kasmi F, Hyunh M, Nishimura EO, Sondek JE, Dangl JL. 2017. TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in Arabidopsis. *Proc Natl Acad Sci U S A* 114:E2053–E2062.

Ntoukakis V, Saur IML, Conlan B, Rathjen JP. 2014. The changing of the guard: the Pto/Prf receptor complex of tomato and pathogen recognition. *Curr Opin Plant Biol* 20C:69–74.

Oda K, Kitano H. 2006. A comprehensive map of the toll-like receptor signaling network. *Mol Syst Biol* 2:2006.0015.

Okuyama Y, Kanzaki H, Abe A, Yoshida K, Tamiru M, Saitoh H, Fujibe T, Matsumura H, Shenton M, Galam DC, Undan J, Ito A, Sone T, Terauchi R. 2011. A multifaceted genomics approach allows the isolation of the rice *Pia*-blast resistance gene consisting of two adjacent NBS-LRR protein genes. *Plant J* 66:467–479.

Ortiz D, De Guillen K, Césari S, Chalvon V, Gracy J, Padilla A, Kroj T. 2017. Recognition of the *Magnaporthe oryzae* effector AVR-Pia by the decoy domain of the rice NLR immune receptor RGA5. *Plant Cell* 29:156–168.

Ortlund E a, Bridgham JT, Redinbo MR, Thornton JW. 2007. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317:1544–1548.

Ose T, Oikawa A, Nakamura Y, Maenaka K, Higuchi Y, Satoh Y, Fujiwara S, Demura M, Sone T, Kamiya M. 2015. Solution structure of an avirulence protein, AVR-Pia, from *Magnaporthe oryzae*. *J Biomol NMR* 63:229–235.

Padmanabhan M, Cournoyer P, Dinesh-Kumar SP. 2009. The leucine-rich repeat domain in plant innate immunity: A wealth of possibilities. *Cell Microbiol* 11:191–198.

Palmer ME, Feldman MW. 2012. Survivability is more fundamental than evolvability. *PLoS One* 7:e38025.

Pandey S, Somssich IE. 2009. The role of WRKY transcription factors in plant immunity. *Plant Physiol* 150:1648–1655.

Parrent JL, James TY, Vasaitis R, Taylor AF. 2009. Friend or foe? Evolutionary history of glycoside hydrolase family 32 genes encoding for sucrolytic activity in fungi and its implications for plant-fungal symbioses. *BMC Evol Biol* 9:148.

Periyannan S, Moore J, Ayliffe M, Bansal U, Wang X, Huang L, Deal K, Luo M, Kong X, Bariana H, Mago R, McIntosh R, Dodds P, Dvorak J, Lagudah E. 2013. The gene *Sr33*, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. *Science* 341:786–788.

Petit-Houdenot Y, Langner T, Harant A, Win J, Kamoun S. 2020. A clone resource of *Magnaporthe oryzae* effectors that share sequence and structural similarities across host-specific lineages. *Mol Plant-Microbe Interact* 33:1032–1035.

Pillai AS, Chandler SA, Liu Y, Signore A V., Cortez-Romero CR, Benesch JLP, Laganowsky A, Storz JF, Hochberg GKA, Thornton JW. 2020. Origin of complexity in haemoglobin evolution. *Nature* 581:480–485.

Potterton L, Agirre J, Ballard C, Cowtan K, Dodson E, Evans PR, Jenkins HT, Keegan R, Krissinel E, Stevenson K, Lebedev A, McNicholas SJ, Nicholls RA, Noble M, Pannu NS, Roth C, Sheldrick G, Skubak P, Turkenburg J, Uski V, von Delft F, Waterman D, Wilson K, Winn M, Wojdyr M. 2018. CCP 4 i 2: the new graphical user interface to the CCP 4 program suite. *Acta Crystallogr Sect D Struct Biol* 74:68–84.

Prigozhin DM, Krasileva KV. 2020. Intraspecies diversity reveals a subset of highly variable plant immune receptors and predicts their binding sites. *bioRxiv* 190785.

Pupko T, Pe'er I, Shamir R, Graur D. 2000. A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol Biol Evol* 17:890–896.

Qi D, DeYoung BJ, Innes RW. 2012. Structure-function analysis of the coiled-coil and leucine-rich repeat domains of the RPS5 disease resistance protein. *Plant Physiol* 158:1819–32.

Qi T, Seong K, Thomazella DPT, Kim JR, Pham J, Seo E, Cho MJ, Schultink A, Staskawicz BJ. 2018. NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in *Nicotiana benthamiana*. *Proc Natl Acad Sci U S A* 115:E10979–E10987.

Qu S, Liu G, Zhou B, Bellizzi M, Zeng L, Dai L, Han B, Wang G-L. 2006. The broad-spectrum blast resistance gene *Pi9* encodes a nucleotide-binding site–leucine-rich repeat protein and is a member of a multigene family in rice. *Genetics* 172:1901–1914.

Radakovic ZS, Anjam MS, Escobar E, Chopra D, Cabrera J, Silva AC, Escobar C, Sobczak M, Grundler FMW, Siddique S. 2018. Arabidopsis HIPP27 is a host susceptibility gene for the beet cyst nematode *Heterodera schachtii*. *Mol Plant Pathol* 19:1917–1928.

Radhakrishnan GV, Keller J, Rich MK, Vernié T, Mbadinga Mbadinga DL, Vigneron N, Cottret L, Clemente HS, Libourel C, Cheema J, Linde AM, Eklund DM, Cheng S, Wong GKS, Lagercrantz U, Li FW, Oldroyd GED, Delaux PM. 2020. An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages.

Nat Plants 6:280–289.

Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, Jiang RHY, Zody MC, Kunjeti SG, Donofrio NM, Meyers BC, Nusbaum C, Kamoun S. 2010. Genome evolution following host jumps in the irish potato famine pathogen lineage. *Science* 330:1540–1543.

Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat Rev Microbiol* 10:417–430.

Rairdan GJ, Collier SM, Sacco M a, Baldwin TT, Boettrich T, Moffett P. 2008. The coiled-coil and nucleotide binding domains of the Potato Rx disease resistance protein function in pathogen recognition and signaling. *Plant Cell* 20:739–751.

Rairdan GJ, Moffett P. 2006. Distinct domains in the ARC region of the potato resistance protein Rx mediate LRR binding and inhibition of activation. *Plant Cell* 18:2082–2093.

Randall RN, Radford CE, Roof KA, Natarajan DK, Gaucher EA. 2016. An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat Commun* 7:1–6.

Ravensdale M, Bernoux M, Ve T, Kobe B, Thrall PH, Ellis JG, Dodds PN. 2012. Intramolecular interaction influences binding of the flax L5 and L6 resistance proteins to their AvrL567 ligands. *PLoS Pathog* 8:e1003004.

Read AC, Hutin M, Moscou MJ, Rinaldi FC, Bogdanove AJ. Cloning of the rice *Xo1* resistance gene and interaction of the Xo1 protein with the defense-suppressing *Xanthomonas* effector Tal2h. *Mol Plant Microbe Interact* 33:1189–1195.

Read AC, Moscou MJ, Zimin A V., Pertea G, Meyer RS, Purugganan MD, Leach JE, Triplett LR, Salzberg SL, Bogdanove AJ. 2020. Genome assembly and characterization of a complex zfBED-NLR gene-containing disease resistance locus in Carolina Gold Select rice with Nanopore sequencing. *PLoS Genet* 16:1–24.

Rice Chromosomes 11 and 12 Sequencing Consortia. 2005. The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. *BMC Biol* 3:20.

Rokas A, Carroll SB. 2008. Frequent and widespread parallel evolution of protein sequences. *Mol Biol Evol* 25:1943–1953.

Rouet R, Langley DB, Schofield P, Christie M, Roome B, Porebski BT, Buckle AM, Clifton BE, Jackson CJ, Stock D, Christ D. 2017. Structural reconstruction of protein ancestry. *Proc Natl Acad Sci U S A* 114:3897–3902.

Sacco MA, Mansoor S, Moffett P. 2007. A RanGAP protein physically interacts with the NB-LRR protein Rx, and is required for Rx-mediated viral resistance. *Plant J* 52:82–93.

Saitou N, Nei M. 1987. The neighbor-joining method : a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.

Sarris PF, Cevik V, Dagdas G, Jones JDG, Krasileva KV. 2016. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biol* 14:8.

Sarris PF, Duxbury Z, Huh SU, Ma Y, Segonzac C, Sklenar J, Derbyshire P, Cevik V, Rallapalli G, Saucet SB, Wirthmueller L, Menke FLH, Sohn KH, Jones JDG. 2015. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell* 161:1089–1100.

Saucet SB, Ma Y, Sarris PF, Furzer OJ, Sohn KH, Jones JDG. 2015. Two linked pairs of Arabidopsis TNL resistance genes independently confer recognition of bacterial effector AvrRps4. *Nat Commun* 6:1–12.

Saur IML, Bauer S, Kracher B, Lu X, Franzeskakis L, Müller MC, Sabelleck B, Kümmel F, Panstruga R, Maekawa T, Schulze-Lefert P. 2019. Multiple pairs of allelic MLA immune receptor-powdery mildew AVR a effectors argue for a direct recognition mechanism. *Elife* 8:1–31.

Schreiber KJ, Bentham A, Williams SJ, Kobe B, Staskawicz BJ. 2016. multiple domain associations within the Arabidopsis immune receptor RPP1 regulate the activation of programmed cell death. *PLOS Pathog* 12:e1005769.

Seeholzer S, Tsuchimatsu T, Jordan T, Bieri S, Pajonk S, Yang W, Jahoor A, Shimizu KK, Keller B, Schulze-Lefert P. 2010. Diversity at the *Mla* powdery mildew resistance locus from cultivated barley reveals sites of positive selection. *Mol Plant-Microbe Interac*t 23:497–509.

Segretin ME, Pais M, Franceschetti M, Chaparro-Garcia A, Bos JIB, Banfield MJ, Kamoun S. 2014. Single amino acid mutations in the potato immune receptor R3a expand response to *Phytophthora* effectors. *Mol Plant-Microbe Interact* 27:624–637.

Seidl MF, Thomma BPHJ. 2017. Transposable elements direct the coevolution between plants and microbes. *Trends Genet* 33:842–851.

Sela H, Spiridon LN, Ashkenazi H, Bhullar NK, Brunner S, Petrescu A-J, Fahima T, Keller B, Jordan T. 2014. Three-dimensional modeling and diversity analysis reveals distinct AVR recognition sites and evolutionary pathways in wild and domesticated wheat *Pm3* R genes. *Mol Plant-Microbe Interact* 27:835–845.

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

Serra H, Choi K, Zhao X, Blackwell AR, Kim J, Henderson IR. 2018. Interhomolog polymorphism shapes meiotic crossover within the Arabidopsis RAC1 and RPP13 disease resistance genes. *PLoS Genet* 14:1–28.

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

Shang J, Tao Y, Chen X, Zou Y, Lei C, Wang J, Li X, Zhao X, Zhang M, Lu Z, Xu J, Cheng Z, Wan J, Zhu L. 2009. Identification of a new rice blast resistance gene, Pid3, by genomewide comparison of paired nucleotide-binding site-leucine-rich repeat genes and their pseudogene alleles between the two sequenced rice genomes. *Genetics* 182:1303–1311.

Shao Z-QQ, Xue J-YY, Wu P, Zhang Y-MM, Wu Y, Hang Y-YY, Wang B, Chen J-QQ. 2016. Large-scale analyses of angiosperm nucleotide-binding site-leucine-rich repeat genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant Physiol* 170:2095–2109.

Shen J, Araki H, Chen L, Chen JQ, Tian D. 2006. Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. *Genetics* 172:1243–1250.

Sinapidou E, Williams K, Nott L, Bahkt S, Tör M, Crute I, Bittner-Eddy P, Beynon J. 2004. Two TIR:NB:LRR genes are required to specify resistance to *Peronospora parasitic*a isolate Cala2 in Arabidopsis. *Plant J* 38:898–909.

Singh R, Dangol S, Chen Y, Choi J, Cho Y-S, Lee J-E, Choi M-O, Jwa N-S. 2016. *Magnaporthe oryzae* effector AVR-Pii helps to establish compatibility by inhibition of the rice NADP-malic enzyme resulting in disruption of oxidative burst and host innate immunity. *Mol Cells* 39:426–438.

Slootweg EJ, Spiridon LN, Roosien J, Butterbach P, Pomp R, Westerhof L, Wilbers R, Bakker E, Bakker J, Petrescu A-J, Smant G, Goverse A. 2013. Structural determinants at the interface of the ARC2 and leucine-rich repeat domains control the activation of the plant immune receptors Rx1 and Gpa2. *Plant Physiol* 162:1510–28.

Smakowska-Luzan E, Mott GA, Parys K, Stegmann M, Howton TC, Layeghifard M, Neuhold J, Lehner A, Kong J, Grünwald K, Weinberger N, Satbhai SB, Mayer D, Busch W, Madalinski M, Stolt-Bergner P, Provart NJ, Mukhtar MS, Zipfel C, Desveaux D, Guttman DS, Belkhadir Y. 2018. An extracellular network of Arabidopsis leucine-rich repeat receptor kinases. *Nature* 553:342–346.

Smith SM, Hulbert SH. 2005. Recombination events generating a novel Rp1 race specificity. *Mol Plant-Microbe Interact* 18:220–228.

Smith SM, Steinau M, Trick HN, Hulbert SH. 2010. Recombinant Rp1 genes confer necrotic or nonspecific resistance phenotypes. *Mol Genet Genomics* 283:591–602.

Snider J, Houry WA. 2008. AAA+ proteins: diversity in function, similarity in structure. *Biochem Soc Trans* 36:72–77.

Staal J, Kaliff M, Dewaele E, Persson M, Dixelius C. 2008. RLM3, a TIR domain encoding gene involved in broad-range immunity of Arabidopsis to necrotrophic fungal pathogens. *Plant J* 55:188–200.

Stam R, Scheikl D, Tellier A. 2016. Pooled enrichment sequencing identifies diversity and evolutionary pressures at NLR resistance genes within a wild tomato population. *Genome Biol Evol* 8:1501–1515.

Stam R, Silva-Arias GA, Tellier A. 2019. Subsets of NLR genes show differential signatures of adaptation during colonization of new habitats. *New Phytol* 224:367–379.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

Starr TN, Picton LK, Thornton JW. 2017. Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* 549:409–413.

Starr TN, Thornton JW. 2016. Epistasis in protein evolution. *Protein Sci* 25:1204–1218.

Steele JFC, Hughes RK, Banfield MJ. 2019. Structural and biochemical studies of an NB-ARC domain from a plant NLR immune receptor. *PLoS One* 14:e0221226.

Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, Chougule K, Gao D, Iwata A, Goicoechea JL, Wei S, Wang J, Liao Y, Wang M, Jacquemin J, Becker C, Kudrna D, Zhang J, Londono CEM, Song X, Lee S, Sanchez P, Zuccolo A, Ammiraju JSS, Talag J, Danowitz A, Rivera LF, Gschwend AR, Noutsos C, Wu C, Kao S, Zeng J, Wei F, Zhao Q, Feng Q, El Baidouri M, Carpentier M-C, Lasserre E, Cooke R, Rosa Farias D da, da Maia LC, dos Santos RS, Nyberg KG, McNally KL, Mauleon R, Alexandrov N, Schmutz J, Flowers D, Fan C, Weigel D, Jena KK, Wicker T, Chen M, Han B, Henry R, Hsing YC, Kurata N, de Oliveira AC, Panaud O, Jackson SA, Machado CA, Sanderson MJ, Long M, Ware D, Wing RA. 2018. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat Genet* 50:285–296.

Steinbrenner AD, Goritschnig S, Krasileva KV., Schreiber KJ, Staskawicz BJ. 2012. effector recognition and activation of the arabidopsis thaliana NLR innate immune receptors. *Cold Spring Harb Symp Quant Biol* 77:249–257.

Steinbrenner AD, Goritschnig S, Staskawicz BJ. 2015. Recognition and activation domains contribute to allele-specific responses of an Arabidopsis NLR receptor to an oomycete effector protein. *PLoS Pathog* 11:1–19.

Steuernagel B, Jupe F, Witek K, Jones JDG, Wulff BBH. 2015. NLR-parser: rapid annotation of plant NLR complements. *Bioinformatics* 31:1665–1667.

Steuernagel B, Periyannan SK, Hernández-Pinzón I, Witek K, Rouse MN, Yu G, Hatta A, Ayliffe M, Bariana H, Jones JDG, Lagudah ES, Wulff BBH. 2016. Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat Biotechnol* 34:652–655.

Steuernagel B, Witek K, Krattinger SG, Ramirez-Gonzalez RH, Schoonbeek HJ, Yu G, Baggs E, Witek AI, Yadav I, Krasileva KV., Jones JDG, Uauy C, Keller B, Ridout CJ, Wulff BBH. 2020. The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. *Plant Physiol 183*:468–482.

Studier FW. 2005. Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif* 41:207–234.

Su YCFF, Bahl J, Joseph U, Butt KM, Peck HA, Koay ESCC, Oon LLEE, Barr IG, Vijaykrishna D, Smith GJDD. 2015. Phylodynamics of H1N1/2009 influenza reveals the transition from host adaptation to immune-driven selection. *Nat Commun* 6:7952.

Sukarta OCA, Slootweg EJ, Goverse A. 2016. Structure-informed insights for NLR functioning in plant immunity. *Semin Cell Dev Biol* 56:134–149.

Sun Y, Li L, Macho AP, Han Z, Hu Z, Zipfel C, Zhou J-M, Chai J. 2013. Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. *Science* 342:624–8.

Takken FLW, Goverse A. 2012. How to build a pathogen detector: structural basis of NB-LRR function. *Curr Opin Plant Biol* 15:375–384.

Talbot NJ. 2003. On the trail of a cereal killer: Exploring the biology of *Magnaporthe grisea*. *Annu Rev Microbiol* 57:177–202.

Tamborski J, Krasileva KV. 2020. Evolution of plant NLRs: from natural history to precise modifications. *Annu Rev Plant Biol* 71:355–378.

Tameling WIL, Baulcombe DC. 2007. Physical association of the NB-LRR resistance protein Rx with a Ran GTPase-activating protein is required for extreme resistance to Potato virus X. *Plant Cell* 19:1682–1694.

Tameling WIL, Elzinga SDJ, Darmin PS, Vossen JH, Takken FLW, Haring MA, Cornelissen BJC. 2002. The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. *Plant Cell* 14:2929–2939.

Tanaka S, Brefort T, Neidig N, Djamei A, Kahnt J, Vermerris W, Koenig S, Feussner K, Feussner I, Kahmann R. 2014. A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize. *Elife* 3:e01355.

Tanaka S, Schweizer G, Rössel N, Fukada F, Thines M, Kahmann R. 2019. Neofunctionalization of the secreted Tin2 effector in the fungal pathogen *Ustilago maydis*. *Nat Microbiol* 4:251–257.

Tarr DEK, Alexander HM. 2009. TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC Res Notes* 2:197.

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences, *American Mathematical Society: Lectures on Mathematics in the Life Sciences.*

Thind AK, Wicker T, Šimková H, Fossati D, Moullet O, Brabant C, Vrána J, Doležel J, Krattinger SG. 2017. Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly. *Nat Biotechnol* 35:793–796.

Thornton JW. 2004. Resurrecting ancient genes: Experimental analysis of extinct molecules. *Nat Rev Genet* 5:366–375.

Todesco M, Kim ST, Chae E, Bomblies K, Zaidem M, Smith LM, Weigel D, Laitinen RAE. 2014. Activation of the *Arabidopsis thaliana* immune system by combinations of common ACD6 alleles. *PLoS Genet* 10: e1004459.

Tokuriki N, Tawfik DS. 2009. Protein dynamism and evolvability. *Science* 324:203–207.

Tosa Y, Osue J, Eto Y, Oh H-S, Nakayashiki H, Mayama S, Leong S a. 2005. Evolution of an avirulence gene, AVR1-CO39, concomitant with the evolution and differentiation of *Magnaporthe oryzae*. *Mol Plant Microbe Interact* 18:1148–60.

Tran DTN, Chung E-H, Habring-Müller A, Demar M, Schwab R, Dangl JL, Weigel D, Chae E. 2017. Activation of a plant NLR complex through heteromeric association with an autoimmune risk variant of another NLR. *Curr Biol* 27:1148–1160.

Ueda H, Yamaguchi Y, Sano H. 2006. Direct interaction between the tobacco mosaic virus helicase domain and the ATP-bound resistance protein, N factor during the hypersensitive response in tobacco plants. *Plant Mol Biol* 61:31–45.

Upson JL, Zess EK, Białas A, Wu C, Kamoun S. 2018. The coming of age of EvoMPMI: evolutionary molecular plant–microbe interactions across multiple timescales. *Curr Opin Plant Biol* 44:108–116.

Urbach JM, Ausubel FM. 2017. The NBS-LRR architectures of plant R-proteins and metazoan NLRs evolved in independent events. *Proc Natl Acad Sci U S A* 114:1063–1068.

Van de Weyer AL, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, Witek K, Jones JDG, Dangl JL, Weigel D, Bemm F. 2019. A species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. *Cell* 178:1260-1272.e14.

van der Hoorn RALL, Kamoun S. 2008. From Guard to Decoy: a new model for perception of plant pathogen effectors. *Plant Cell* 20:2009–17.

Van Der Vossen EAG, Van Der Voort JNAMR, Kanyuka K, Bendahmane A, Sandbrink H, Baulcombe DC, Bakker J, Stiekema WJ, Klein-Lankhorst RM. 2000. Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: A virus and a nematode. *Plant J* 23:567–576.

van Wersch S, Li X. 2019. Stronger when together: clustering of plant NLR disease resistance genes. T*rends Plant Sci* 24:688–699.

Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ. 2012. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol* 10:e1001446.

Wan L, Essuman K, Anderson RG, Sasaki Y, Monteiro F, Chung E-H, Osborne Nishimura E, DiAntonio A, Milbrandt J, Dangl JL, Nishimura MT. 2019. TIR domains of plant immune receptors are NAD + -cleaving enzymes that promote cell death. *Science* 365:799–803.

Wang G-F, Ji J, EI-Kasmi F, Dangl JL, Johal G, Balint-Kurti PJ. 2015. Molecular and functional analyses of a maize autoactive NB-LRR protein identify precise structural requirements for activity. *PLOS Pathog* 11:e1004674.

Wang G, Roux B, Feng F, Guy E, Li L, Li N, Zhang X, Lautier M, Jardinaud M-FF, Chabannes M, Arlat M, Chen S, He C, Noël LD, Zhou J-MM, Noel LD, Zhou J-MM. 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 H, Sun S, Ge W, Zhao L, Hou B, Wang K, Lyu Z, Chen L, Xu S, Guo J, Li M, Su P, Li Xuefeng, Wang G, Bo C, Fang X, Zhuang W, Cheng X, Wu J, Dong L, Chen W, Li W, Xiao G, Zhao J, Hao Y, Xu Y, Gao Y, Liu W, Liu Y, Yin H, Li J, Li Xiang, Zhao Y, Wang X, Ni F, Ma X, Li A, Xu SS, Bai G, Nevo E, Gao C, Ohm H, Kong L. 2020. Horizontal gene transfer of Fhb7 from fungus underlies *Fusarium* head blight resistance in wheat. *Science* 368:eaba5435.

Wang J, Wang Jia, Hu M, Wu S, Qi J, Wang G, Han Z, Qi Y, Gao N, Wang H-W, Zhou J-M, Chai J. 2019. Ligand-triggered allosteric ADP release primes a plant NLR complex. *Science* 364:eaav5868.

Wang JJ, Hu M, Wang JJ, Qi J, Han Z, Wang G, Qi Y, Wang H-W, Zhou J-M, Chai J. 2019. Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science* 364:eaav5870.

Wang L, Xu X, Lin F, Pan Q. 2009. Characterization of rice blast resistance genes in the *Pik* cluster and fine mapping of the *Pik-p* locus. *Phytopathology* 99:900–905.

Wang Z-X, Yano M, Yamanouchi U, Iwamoto M, Monna L, Hayasaka H, Katayose Y, Sasaki T. 1999. The P*ib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. *Plant J* 19:55–64.

Washburn JD, Bird KA, Conant GC, Pires JC. 2016. Convergent evolution and the origin of complex phenotypes in the age of systems biology. *Int J Plant Sci* 177:305–318.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46:W296–W303.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191.

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

Wei F, Wing R a, Wise RP. 2002. Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. *Plant Cell* 14:1903–1917.

Wei T, Gong J, Jamitzky F, Heckl WM, Stark RW, Rössle SC. 2008. LRRML: a conformational database and an XML description of leucine-rich repeats (LRRs). *BMC Struct Biol* 8:47.

Weinreich DM, Delaney Nigel F, Depristo Mark A, Hartl Daniel L, Gross R, Yelin I, Kishony Roy, Weinreich DM, Delaney N. F., Depristo M. A., Hartl D. L., Lee HH, Molla MN, Cantor CR, Collins JJ, Lane PG, Hutter A, Oliver SG, Butler PR, Toprak E, Veres A, Michel JB, Chait R, Hartl D. L., Kishony R., Imamovic L, Sommer MOA, Lázár V, Singh GP, Spohn R, Nagy I, Horváth B, Hrtyan M, Busa-Fekete R, Bogos B, Méhi O, Csörgő B, Pósfai G, Fekete G, Szappanos B, Kégl B, Papp B, Pál C, Fridman O, Goldberg A, Ronin I, Shoresh N, Balaban NQ, Hegreness M, Shoresh N, Damian D, Hartl D, Kishony R., Zhang Q, Lambert G, Liao D, Kim H, Robin K, Tung CK, Pourmand N, Austin RH, Conrad TM, Lewis NE, Palsson BO, Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC, Martínez JL, Hallatschek O, Hersen P, Ramanathan S, Nelson DR, Greulich P, Waclaw B, Allen RJ, Hermsen R, Deris JB, Hwa T, Berg HC, Brown DA, Xavier JB, Foster KR, Bell G, Gonzalez A, Gonzalez A, Bell G, Lindsey HA, Gallie J, Taylor S, Kerr B, Bridle JR, Vines TH, Polechová J, Barton NH, Fowler RG, Degnen GE, Cox EC, Echols H, Lu C, Burgers PM, Galán J-C, Tato M, Baquero MR, Turrientes C, Baquero F, Martinez JL, Smith DR, Calvo JM, Martin RG, Jair KW, Wolf RE, Rosner JL, Bollenbach T, Quan S, Chait R, Kishony R., Lenski RE, Levin BR, Perrot V, Walker

N, Handel A, Regoes RR, Antia R, Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C, Sanjuán R, Cuevas JM, Furió V, Holmes EC, Moya A, Codoñer FM, Darós J-A, Solé R V., Elena SF, Muller HJ, Barrick JE, Lenski RE, Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muñiz-Rascado L, Ong Q, Paley S, Schröder I, Shearer AG, Subhraveti P, Travers M, Weerasinghe D, Weiss V, Collado-Vides J, Gunsalus RP, Paulsen I, Karp PD, Baba T, Ara T, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H, Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R., Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW, Langmead B, Salzberg SL, DePristo MA, Banks E, Poplin R, Garimella K V., Maguire JR, Hartl C, Philippakis AA, Angel G del, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312:111–114.

Wendler P, Ciniawsky S, Kock M, Kube S. 2012. Structure and function of the AAA+ nucleotide binding pocket. *Biochim Biophys Acta - Mol Cell Res* 1823:2–14.

Wicker T, Buchmann JP, Keller B. 2010. Patching gaps in plant genomes results in gene movement and erosion of colinearity. *Genome Res* 20:1229–1237.

Wicker T, Yahiaoui N, Keller B. 2007. Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes. *Plant J* 51:631–641.

Williams SJ, Sohn KH, Wan L, Bernoux M, Sarris PF, Segonzac C, Ve T, Ma Y, Saucet SB, Ericsson DJ, Casey LW, Lonhienne T, Winzor DJ, Zhang X, Coerdt A, Parker JE, Dodds PN, Kobe B, Jones JDG. 2014. Structural basis for assembly and function of a heterodimeric plant immune receptor. *Science* 344:299–303.

Williams SJ, Sornaraj P, DeCourcy-Ireland E, Menz RI, Kobe B, Ellis JG, Dodds PN, Anderson P a. 2011. An autoactive mutant of the M flax rust resistance protein has a preference for binding ATP, whereas wild type m protein binds ADP. *Mol Plant-Microbe Interact* 24:897–906.

Win J, Chaparro-Garcia A, Belhaj K, Saunders DGO, Yoshida K, Dong S, Schornack S, Zipfel C, Robatzek S, Hogenhout SA, Kamoun S. 2012. Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb Symp Quant Biol* 77:235–47.

Win J, Kamoun S. 2003. pCB301-p19: A binary plasmid vector to enhance transient expression of transgenes by agroinfiltration. *Plant J* 33:949–956.

Win J, Kamoun S, Jones AME. 2011. Purification of Effector–Target Protein Complexes via Transient Expression in *Nicotiana benthamiana*, Plant Immunity, *Methods in Molecular Biology*. Totowa, NJ: Humana Press.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP 4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* 67:235–242.

Winter G. 2010. xia2 : an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* 43:186–190.

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

Wu C-H, Belhaj K, Bozkurt TO, Kamoun S. 2016. The NLR helper protein NRC3 but not NRC1 is required for Pto-mediated cell death in *Nicotiana benthamiana*. *New Phytol* 209:1344–1352.

Wu C-H, Krasileva KV., Banfield MJ, Terauchi R, Kamoun S. 2015. The 'sensor domains' of plant NLR proteins: more than decoys? *Front Plant Sci* 6:5–7.

Wu J, Kou Y, Bao J, Li Y, Tang M, Zhu X, Ponaya A, Xiao G, Li J, Li C, Song MY, Cumagun CJR, Deng Q, Lu G, Jeon JS, Naqvi NI, Zhou B. 2015. Comparative genomics identifies the *Magnaporthe oryzae* avirulence effector AvrPi9 that triggers Pi9-mediated blast resistance in rice. *New Phytol* 206:1463–1475.

Wu Z, Li M, Dong OX, Xia S, Liang W, Bao Y, Wasteneys G, Li X. 2019. Differential regulation of TNL-mediated immune signaling by redundant helper CNLs. *New Phytol* 222:938–953.

Xiahong HE, Yan SUN, Dong GAO, Fugang WEI, Lei PAN, Cunwu GUO, Ruzhi MAO, Yong XIE, Chengyun LI, Youyong ZHU. 2011. Comparison of agronomic traits between rice landraces and modern varieties at different altitudes in the paddy fields of Yuanyang Terrace, Yunnan Province. *J Resour Ecol* 2:46–50.

Xiong Y, Han Z, Chai J. 2020. Resistosome and inflammasome: platforms mediating innate immunity. *Curr Opin Plant Biol* 56:47–55.

Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J, Orjeda G, Guzman F, Torres M, Lozano R, Ponce O, Martinez D, De La Cruz G, Chakrabarti SK, Patil VU, Skryabin G, Kuznetsov BB, Ravin N V., Kolganova T V., Beletsky A V., Mardanov A V., Di Genova A, Bolser DM, Martin DMA, Li G, Yang Y, Kuang H, Hu Q, Xiong X, Bishop GJ, Sagredo B, Mejía N, Zagorski W, Gromadka R, Gawor J, Szczesny P, Huang S, Zhang Z, Liang C, He J, Li Y, He Y, Xu J, Zhang Y, Xie B, Du Y, Qu D, Bonierbale M, Ghislain M, Herrera MDR, Giuliano G, Pietrella M, Perrotta G, Facella P, O'Brien K, Feingold SE, Barreiro LE, Massa GA, Diambra L, Whitty BR, Vaillancourt B, Lin H, Massa AN, Geoffroy M, Lundback S, DellaPenna D, Buell CR, Sharma SK, Marshall DF, Waugh R, Bryan GJ, Destefanis M, Nagy I, Milbourne D, Thomson SJ, Fiers M, Jacobs JME, Nielsen KL, Sønderkær M, Iovene M, Torres GA, Jiang J, Veilleux RE, Bachem CWB, De Boer J, Borm T, Kloosterman B, Van Eck H, Datema E, Hekkert BL, Goverse A, Van Ham RCHJ, Visser RGF. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475:189–195.

Yamamoto E, Takashi T, Morinaka Y, Lin S, Wu J, Matsumoto T, Kitano H, Matsuoka M, Ashikari M. 2010. Gain of deleterious function causes an autoimmune response and Bateson-Dobzhansky-Muller incompatibility in rice. *Mol Genet Genomics* 283:305–315.

Yang Z. 2005. Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Mol Biol Evol* 22:1107–1118.

Yang Z. 2000. maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J Mol Evol* 51:423–432.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* 13:555–556.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* 17:32–43.

Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol* 46:409–418.

Yang Z, Swanson WJ, Vacquier VD. 2000. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol Biol Evol* 17:1446–1455.

Yoshida K, Saunders DGOO, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, Inoue Y, Chuma I, Tosa Y, Cano LM, Kamoun S, Terauchi R. 2016. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. *BMC Genomics* 17:370.

Yoshida Kentaro, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Yoshida Kakoto, Tosa Y, Chuma I, Takano Y, Win J, Kamoun S, Terauchi R. 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* 21:1573–91.

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang Z-X, Kono I, Kurata N, Yano M, Iwata N, Sasaki T. 1998. Expression of Xa1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proc Natl Acad Sci U S A* 95:1663–1668.

Yuan B, Zhai C, Wang W, Zeng X, Xu X, Hu H, Lin F, Wang L, Pan Q. 2011. The Pik-p resistance to *Magnaporthe oryzae* in rice is mediated by a pair of closely linked CC-NBS-LRR genes. *Theor Appl Genet* 122:1017–1028.

Yue JX, Meyers BC, Chen JQ, Tian D, Yang S. 2012. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *New Phytol* 193:1049–1063.

Zdrzałek R, Kamoun S, Terauchi R, Saitoh H, Banfield MJ. 2020. The rice NLR pair Pikp-1/Pikp-2 initiates cell death through receptor cooperation rather than negative regulation. *PLoS One* 15:1–15.

Zess EK, Białas A, Kamoun S. 2019. Old fungus, new trick. Nat Microbiol 4:210–211.

Zhai C, Lin F, Dong Z, He X, Yuan B, Zeng X, Wang L, Pan Q. 2011. The isolation and characterization of Pik, a rice blast resistance gene which emerged after rice domestication. *New Phytol* 189:321–334.

Zhai C, Zhang Y, Yao N, Lin F, Liu Z, Dong Z, Wang L, Pan Q. 2014. Function and interaction of the coupled genes responsible for Pik-h encoded rice blast resistance. *PLoS One* 9:e98067.

Zhang H, Zheng X, Zhang Z. 2016. The *Magnaporthe grisea* species complex and plant pathogenesis. *Mol Plant Pathol* 7:796–804.

Zhang L, Chen S, Ruan J, Wu J, Tong AB, Yin Q, Li Y, David L, Lu A, Wang WL, Marks C, Ouyang Q, Zhang X, Mao Y, Wu H. 2015. Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science* 350:404–409.

Zhang X, Bernoux M, Bentham AR, Newman TE, Ve T, Casey LW, Raaymakers TM, Hu J, Croll TI, Schreiber KJ, Staskawicz BJ, Anderson PA, Sohn KH, Williams SJ, Dodds PN, Kobe B. 2017. Multiple functional self-association interfaces in plant TIR domains. *Proc Natl Acad Sci U S A* 114:E2046–E2052.

Zhang Y, Xia R, Kuang H, Meyers BC. 2016. The Diversification of Plant NBS-LRR Defense Genes Directs the Evolution of MicroRNAs That Target Them. Mol Biol Evol 33:2692–2705.

Zhou B, Qu S, Liu G, Dolan M, Sakai H, Lu G, Bellizzi M, Wang G-L. 2006. The eight amino-acid differences within three leucine-rich repeats between Pi2 and Piz-t resistance proteins determine the resistance specificity to *Magnaporthe grisea*. *Mol Plant-Microbe Interact* 19:1216–1228.

Zhou M, Li Y, Hu Q, Bai XC, Huang W, Yan C, Scheres SHW, Shi Y. 2015. Atomic structure of the apoptosome: Mechanism of cytochrome c- and dATP-mediated activation of Apaf-1. *Genes Dev* 29:2349–2361.

Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D. 2004. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol Genet Genomics* 271:402–415.

Zipfel C. 2014. Plant pattern-recognition receptors. Trends Immunol 35:345–351.

Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JDG, Felix G, Boller T. 2004. Bacterial disease resistance in Arabidopsis through flagellin perception. *Nature* 428:764–767.

Zou Z, Zhang J. 2015. Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Mol Biol Evol* 32:2085–2096.

Zschiesche W, Barth O, Daniel K, Böhme S, Rausche J, Humbeck K. 2015. The zinc-binding nuclear protein HIPP3 acts as an upstream regulator of the salicylate-dependent plant immunity pathway and of flowering time in *Arabidopsis thaliana*. *New Phytol* 207:1084–1096.