

RESEARCH ARTICLE

A single amino acid polymorphism in a conserved effector of the multihost blast fungus pathogen expands host-target binding spectrum

Adam R. Bentham¹, Yohann Petit-Houdenot^{2,3}, Joe Win², Izumi Chuma⁴, Ryohei Terauchi^{5,6}, Mark J. Banfield^{1*}, Sophien Kamoun^{2*}, Thorsten Langner^{2*}

1 Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich, United Kingdom, **2** The Sainsbury Laboratory, University of East Anglia, Norwich Research Park, Norwich, United Kingdom, **3** Université Paris-Saclay, INRAE, AgroParisTech, UMR BIOGER, Thiverval-Grignon, France, **4** Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Japan, **5** Kyoto University, Kyoto, Japan, **6** Iwate Biotechnology Research Center, Kitakami, Japan

* Mark.Banfield@jic.ac.uk (MJB); Sophien.Kamoun@tsl.ac.uk (SK); Thorsten.Langner@tsl.ac.uk (TL)



OPEN ACCESS

Citation: Bentham AR, Petit-Houdenot Y, Win J, Chuma I, Terauchi R, Banfield MJ, et al. (2021) A single amino acid polymorphism in a conserved effector of the multihost blast fungus pathogen expands host-target binding spectrum. *PLoS Pathog* 17(11): e1009957. <https://doi.org/10.1371/journal.ppat.1009957>

Editor: Jin-Rong Xu, Purdue University, UNITED STATES

Received: March 29, 2021

Accepted: September 14, 2021

Published: November 10, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.ppat.1009957>

Copyright: © 2021 Bentham et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The authors confirm that all data underlying the findings are fully available without restriction. Sequence information

Abstract

Accelerated gene evolution is a hallmark of pathogen adaptation and specialization following host-jumps. However, the molecular processes associated with adaptive evolution between host-specific lineages of a multihost plant pathogen remain poorly understood. In the blast fungus *Magnaporthe oryzae* (Syn. *Pyricularia oryzae*), host specialization on different grass hosts is generally associated with dynamic patterns of gain and loss of virulence effector genes that tend to define the distinct genetic lineages of this pathogen. Here, we unravelled the biochemical and structural basis of adaptive evolution of APiKL2, an exceptionally conserved paralog of the well-studied rice-lineage specific effector AVR-Pik. Whereas AVR-Pik and other members of the six-gene AVR-Pik family show specific patterns of presence/absence polymorphisms between grass-specific lineages of *M. oryzae*, APiKL2 stands out by being ubiquitously present in all blast fungus lineages from 13 different host species. Using biochemical, biophysical and structural biology methods, we show that a single aspartate to asparagine polymorphism expands the binding spectrum of APiKL2 to host proteins of the heavy-metal associated (HMA) domain family. This mutation maps to one of the APiKL2-HMA binding interfaces and contributes to an altered hydrogen-bonding network. By combining phylogenetic ancestral reconstruction with an analysis of the structural consequences of allelic diversification, we revealed a common mechanism of effector specialization in the AVR-Pik/APiKL2 family that involves two major HMA-binding interfaces. Together, our findings provide a detailed molecular evolution and structural biology framework for diversification and adaptation of a fungal pathogen effector family following host-jumps.

of all isolates used in this study are included in [S1 Table](#). Protein structures, and the data used to derive these, have been deposited at the Protein DataBank (PDB) with accession codes 7NLJ (APikL2A/sHMA25) and 7NMM (APikL2F/sHMA94). The full yeast two-hybrid HMA-domain library is available through Addgene (https://www.addgene.org/Sophien_Kamoun, plasmid accession numbers 168306-168456) and detailed information is available in Langner et al., (2021, Zenodo; <https://zenodo.org/record/5148559#.YQQabC1Q1Xs>).

Funding: Y.P.-H. was supported by a young scientist contract from INRA. S.K. is funded by the Gatsby Charitable Foundation (<https://www.gatsby.org.uk/>) and UK Research and Innovation Biotechnology and Biological Sciences Research Council (UKRI-BBSRC) grant BB/P012574. M.J.B. is funded by the John Innes Foundation (<https://www.johninnesfoundation.org.uk/>). S.K. and M.J.B. receive funding from UKRI-BBSRC grants BBS/E/J/000PR9797 and BBS/E/J/000PR9798, and the European Research Council (ERC) proposal 743165. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Author summary

Plant pathogens secrete effector proteins inside the host where they interact with host target proteins to alter cellular processes and enable infection and disease progression. These effector proteins evolve rapidly to adapt to changing host targets and to avoid recognition by the plant immune system, especially in multihost pathogens that frequently undergo host jumps or host-range expansions. The blast fungus *Magnaporthe oryzae* is one of the most destructive disease of cereals. Host specialization of *M. oryzae* lineages is generally associated with gains and losses of effector genes and selection of adaptive mutations. However, the molecular mechanisms of how effector proteins adapt to host targets is poorly understood. Here, we unravelled the biochemical and structural basis underlying adaptive evolution of the effector APikL2. APikL2 belongs to a six gene family of sequence related effectors including the well-studied effector AVR-Pik. APikL2 is exceptionally conserved across host-specific lineages but displays host-specific polymorphisms. We combine biochemical, structural and evolutionary biology methods to provide a molecular evolutionary framework how this effector family adapts to changing host targets. In APikL2, a single amino acid change expands binding to a host-target of the heavy-metal associated domain containing protein family. By reconstructing the evolutionary history of the effector family, we show that this amino acid polymorphism is a derived trait. We combine the knowledge gained from our evolutionary and structural biology analyses to unravel a common mechanism of adaptation in the APikL effector family that involves diversification of two major target-binding interfaces. Together, our findings expand our knowledge about how pathogen effectors adapt to changing host targets following host jumps.

Introduction

A number of plant pathogens evolve by occasionally shifting or jumping from one host plant to another. Such host jump events are expected to have a dramatic impact on the evolution of pathogen virulence effectors—secreted proteins that can translocate inside host plant cells to modulate host processes, suppress immunity and facilitate colonization [1–3]. As effectors become exposed to a different cellular environment, their evolution is driven by two broad selection pressures imposed by the new host plant [4]. Effectors can evolve to adapt to the cellular targets of their new hosts [1,5] or to evade detection by plant immune receptors that can detect them by direct binding or indirectly through their perturbation of host processes [6–10]. These co-evolutionary conflicts have generally led to expansive and highly diverse effector protein repertoires in plant pathogens, notably in fungal and oomycete pathogens (filamentous pathogens). Effector genes are often located in specific genomic regions that allow for rapid evolution, such as gene gain and loss, emergence of adaptive mutations and horizontal gene transfer [11–14]. In biotrophic filamentous plant pathogens, i.e. pathogens that infect and proliferate in living host tissue, accelerated gene evolution is mostly documented at the level of coevolution with a single host species, but it has also been associated with host-range expansions and host jumps [1,14–21]. However, the molecular and biochemical processes associated with adaptive evolution between host-specific lineages of a multihost plant pathogen remain poorly understood.

The blast fungus *Magnaporthe oryzae* (Syn. *Pyricularia oryzae*) is a plant pathogen best known as the causal agent of rice blast disease. However, despite its Linnean name, *M. oryzae* has an extensive host range comprising more than 50 grass hosts, including major cereal crops

such as wheat and millet as well as their wild relatives (<https://ars-grin.gov/fungalatabases/>). Although *M. oryzae* as a species has dozens of host species, individual isolates tend to preferentially associate with a single grass genus, leading to genetically divergent, host-specific lineages [14,22,23]. The current view is that *M. oryzae* is an assemblage of genetic lineages under incipient speciation following host shifts or host range expansions [23–25]. Extensive genomic resources, including chromosome quality reference genomes of host-specific isolates, have enabled comparative genomics studies and make *M. oryzae* an excellent experimental system to investigate the link between effector diversity and host-specificity [26–28]. To date, over 20 *M. oryzae* effector genes have been experimentally validated, primarily for their avirulence (AVR) activities (activation of immunity), and large-scale effector resources are available [29]. The majority of the studied effector genes show presence/absence (P/A) polymorphisms as a consequence of gene gain/loss associated with colonization of particular host genera or species [14,22,30,31]. In addition, an excess of nonsynonymous substitutions, consistent with positive selection, has been reported for the effector genes AVR-Pik, AVR-Pizt, Slp1, PWL3, AVR-Pita3 and BAS2 [2,14], particularly within the rice-infecting lineage. These signatures of positive selection are generally viewed as adaptive and attributed to arms race interplay with the host plant immune system following the gene for gene model. Nonetheless, phylogenomic analyses that map key biochemical transitions of effector proteins onto a molecular evolution framework remain limited, especially at the species-wide level.

Filamentous pathogen effectors generally expand through duplication and diversification from common ancestors, and therefore tend to share similar features such as conserved structural folds [32]. In *M. oryzae*, a suite of secreted proteins known as MAX (*Magnaporthe oryzae* avirulence and Tox-B-like) effectors groups sequence-unrelated proteins that share a characteristic six-stranded beta-sandwich fold [33]. MAX effectors have expanded in *M. oryzae* accounting for ~5–10% of the total putative effector repertoire and about half of the experimentally validated effectors. Although ~100 MAX fold proteins have been identified by Hidden Markov Model (HMM) predictions [29,33], AVR-Pik, AVR1-CO39, AVR-Pib, AVRPiz-t and AVR-Pia remain the only *M. oryzae* MAX effectors that have been structurally and functionally characterised [8,10,14,34–38]. Three of these MAX effectors (AVR-Pik, AVR1-CO39 and AVR-Pia) bind host proteins carrying a heavy metal-associated (HMA) domain (see below) [10,34,36]. However, it remains unclear whether the MAX fold is predictive of the precise effector activities or whether it serves as a basic structural scaffold that confers protein structural integrity while providing a framework for sequence and functional diversification [14,33].

Although *M. oryzae* effectors have evolved to modulate host processes to benefit the pathogen, they have been initially discovered through their AVR activities or the capacity to activate immune receptors (encoded by resistance genes (R-genes) commonly called Pi genes) in resistant host genotypes. For example, the AVR-Pik and AVR1-CO39/AVR-Pia effectors activate the rice Pik and Pia immune receptor proteins, respectively, granting resistance to *M. oryzae* isolates carrying the cognate effector [39–42]. Both the Pik and Pia loci consist of paired intracellular sensor/helper receptors of the NLR (nucleotide-binding and leucine-rich repeat) class. Remarkably, both sensor NLRs (Pik-1 and Pia-2 syn. RGA5) carry an extraneous integrated domain with similarity to the heavy metal-associated (HMA) proteins that are targeted by the effectors [10,34,36]. These integrated HMA domains of Pik-1 and Pia2 directly bind AVR-Pik and AVR-Pia/AVR1-CO39, respectively. Extensive structure-function studies of the integrated HMA domains in complex with the effectors demonstrated the importance of the binding affinity to the HMA domain for the activation of the NLR proteins [7,10,34,36,43,44]. Interestingly, comparative analyses of the crystal structures of AVR-Pik, AVR-Pia and AVR1-CO39 in complex with HMA domains revealed common and contrasting binding interfaces. Whereas

AVR1-CO39 and AVR-Pia bind to the Pia2-HMA at a single interface formed by hydrophobic interactions between residues of the β -2 strands of both the effector and the HMA [35], AVR-Pik binds to the Pik-1 HMA domain mainly by electrostatic interactions mediated by three distinct interfaces [7,10,43]. The AVR-Pik/Pik-1 HMA interfaces are comprised of main chain interactions between the residues of the C-terminal β -4 strand of the HMA and β -3 of the effector (interface 1), a salt bridge between the β -2 loop of the HMA and β -2 strand of the effector (interface 2), and interactions between the residues of the N-terminal loop of the effector and the β -2 and β -3 strands of the HMA (interface 3).

The remarkable finding that a single type of NLR-integrated domain mediates perception of three sequence diverse blast fungus effectors indicates that this pathogen has evolved multiple effectors to target HMA-containing proteins. Indeed, two recent studies validated the view that, before being baited by NLR proteins, AVR-Pik and other HMA binding effectors have convergently evolved to target host proteins of the heavy metal-associated plant proteins (HPPs) and heavy metal-associated isoprenylated plant proteins (HIPPs), collectively referred to as small HMA (sHMA) proteins [45–47]. sHMA proteins are metallochaperone-like proteins that belong to an expanded family of up to 100 members in a single grass species that contribute to metal homeostasis and detoxification during abiotic stress [45,48]. The exact role of sHMA proteins during plant-pathogen interactions is still unknown and is complicated by their genetic redundancy and complex interplay with multiple pathogen effectors. However, recurrent integration of HMA-domains into the canonical architecture of rice NLR proteins indicates that sHMA proteins are the operative targets of a number of *M. oryzae* MAX effector proteins. The relevance of sHMA proteins during pathogen infection is further supported by the finding that the rice sHMA proteins Pi21 and OsHIP20 are susceptibility factors that supports infection by *M. oryzae* [47,49].

The molecular processes associated with adaptive evolution in host-specific lineages of a multi-host plant pathogen are poorly understood. Although there is a wealth of knowledge about genome and effector gene evolution in the context of gene-for-gene co-evolution with plant immune receptors, there are only a few documented examples of pathogen effector adaptation to changing host target environments, particularly following host jumps and host shifts. Here, we investigated the biochemical and structural basis of adaptive evolution of APikL2 (syn. Pex75 as per [8,42]), an exceptionally conserved paralog of the well-studied effector AVR-Pik that is specific to the rice-lineage of *M. oryzae*. APikL2 stands out compared to the five other members of the AVR-Pik family by having limited presence/absence polymorphisms and being present in all *M. oryzae* lineages from 13 different host species. We combined biochemical, biophysical and structural biology methods to show that a single amino acid polymorphism expands the binding spectrum of APikL2 to a host sHMA protein. This mutation maps to an APikL2-HMA binding interface and contributes to an altered hydrogen-bonding network that may facilitate binding. Further, phylogenetic ancestral reconstruction and analysis of the structural consequences of allelic diversification of the AVR-Pik/APikL2 family revealed that two major HMA-binding interfaces have been targeted by positive selection. Altogether, our findings provide a detailed molecular evolution and structural biology framework for diversification and adaptation of a fungal pathogen effector during adaptation to a new host environment.

Results

A large family of effectors with sequence similarity to AVR-Pik occurs in multiple host-specific lineages of *Magnaporthe oryzae*

To identify AVR-Pik like (APikL) effector candidates across multiple host-specific lineages, we took advantage of 107 publicly available genome assemblies of *M. oryzae*, representative of ten

distinct genetic lineages including isolates from 13 host species (S1 Table). We also included six *Magnaporthe grisea* crabgrass (*Digitaria sanguinalis*) isolates and one fountain grass (*Pennisetum americanum*) isolate as outgroups in the analysis. The final set of *M. oryzae* genomes contained 24 *Oryza sativa*, six *Setaria* spp., nine *Eleusine* spp., 17 *Lolium* spp., 37 *Triticum* spp., one *Eragrostis curvala*, two *Brachiaria* spp., two *Panicum repens*, and two *Stenotaphrum secundatum* infecting isolates from different geographical regions. To test whether these assemblies are suitable for presence/absence analyses, we performed a BUSCO analysis that confirmed completeness of all assemblies at >90% except the *M. grisea* isolates VO107 and Dig41 (BUSCO completeness of 84.4% and 88.7%, respectively).

We then identified APiKL family members using an iterative TBLASTN approach. First, we used AVR-PikD (D variant of AVR-Pik; EnsemblFungi accession number in 70–15 assembly MG08: MGG_15972) as a query to identify sequence related proteins. This led to the identification of four effector candidates that we named APiKL1, APiKL2 (Syn. Pex75; accession number in assembly MQUD00000000.1 of isolate B71: BSY92_00423 according to [28]), APiKL3, and APiKL4. We then repeated the search using all APiKL family members leading to the identification of APiKL5, which is more divergent and carries a 19 amino acid C-terminal extension. In total, we identified 179 genes belonging to a family of six APiKL effectors (including AVR-Pik; S1 File). Ten of these genes appeared to be pseudogenized in isolates belonging to the *Oryza*, *Setaria* and *Brachiaria* lineage 2. We defined genes as pseudogenes when one of the following conditions were met: i) mutated or missing start codon, ii) truncated sequence in the assembly or iii) premature stop codons present. The identified pseudogenes included two AVR-Pik genes that were C-terminally truncated in *Oryza* infecting isolates, one APiKL5 gene that is N-terminally truncated in the *Brachiaria* infecting isolate Bd8401, four APiKL4 genes in *Setaria* infecting isolates in which the start codon was mutated, and three APiKL2 genes with premature stop codons in rice infecting isolates from the US. The exact same mutations were shared between all APiKL4 or APiKL2 pseudogenes, respectively. Commonly, isolates have one or two genes (or pseudogenes) of the family except isolate Bm88324 (*Brachiaria* lineage 1) and four *M. grisea* isolates isolated from *D. sanguinalis* that have no detectable APiKL homologs.

Next, we mapped all APiKL family members onto the genetic lineages of *M. oryzae*. We first generated maximum likelihood trees of 1920 conserved single copy orthologs of all isolates and applied a multispecies coalescence method using ASTRAL [50–52] to infer a coalescence species tree (Fig 1). The topology of the species tree was consistent with previously published analyses [23] and reflected the host-driven genetic specialization of *M. oryzae* lineages.

We then annotated presence/absence of each APiKL family effector across the tree. Notably, most APiKL family members showed lineage-specific presence/absence polymorphisms, similar to other host-specific AVR-effector proteins, with the following specificities: AVR-Pik was found in the *Oryza*-infecting lineage (and the *Panicum repens* isolates BTTrp-5 and -6), APiKL4 in *Setaria*- and *Eleusine*-infecting lineages, and APiKL5 in the *Lolium*-, *Triticum*- and *Eleusine*-infecting *M. oryzae* lineages (Fig 1). APiKL1 and APiKL3 were specific to the *Digitaria*- and *Pennisetum*-infecting species *M. grisea* and *M. pennisetigena*, respectively. Unlike the other APiKL effectors, APiKL2 was exceptionally conserved across all lineages irrespective of the host plant, with the exception of the *M. grisea* and *M. pennisetigena* isolates that represent the outgroup in our analysis.

APiKL2 is exceptionally conserved across all *M. oryzae* lineages but displays lineage-specific polymorphisms

In addition to presence/absence polymorphisms, allelic diversification is often associated with pathogen effectors. To determine the extent of host-lineage specific amino acid

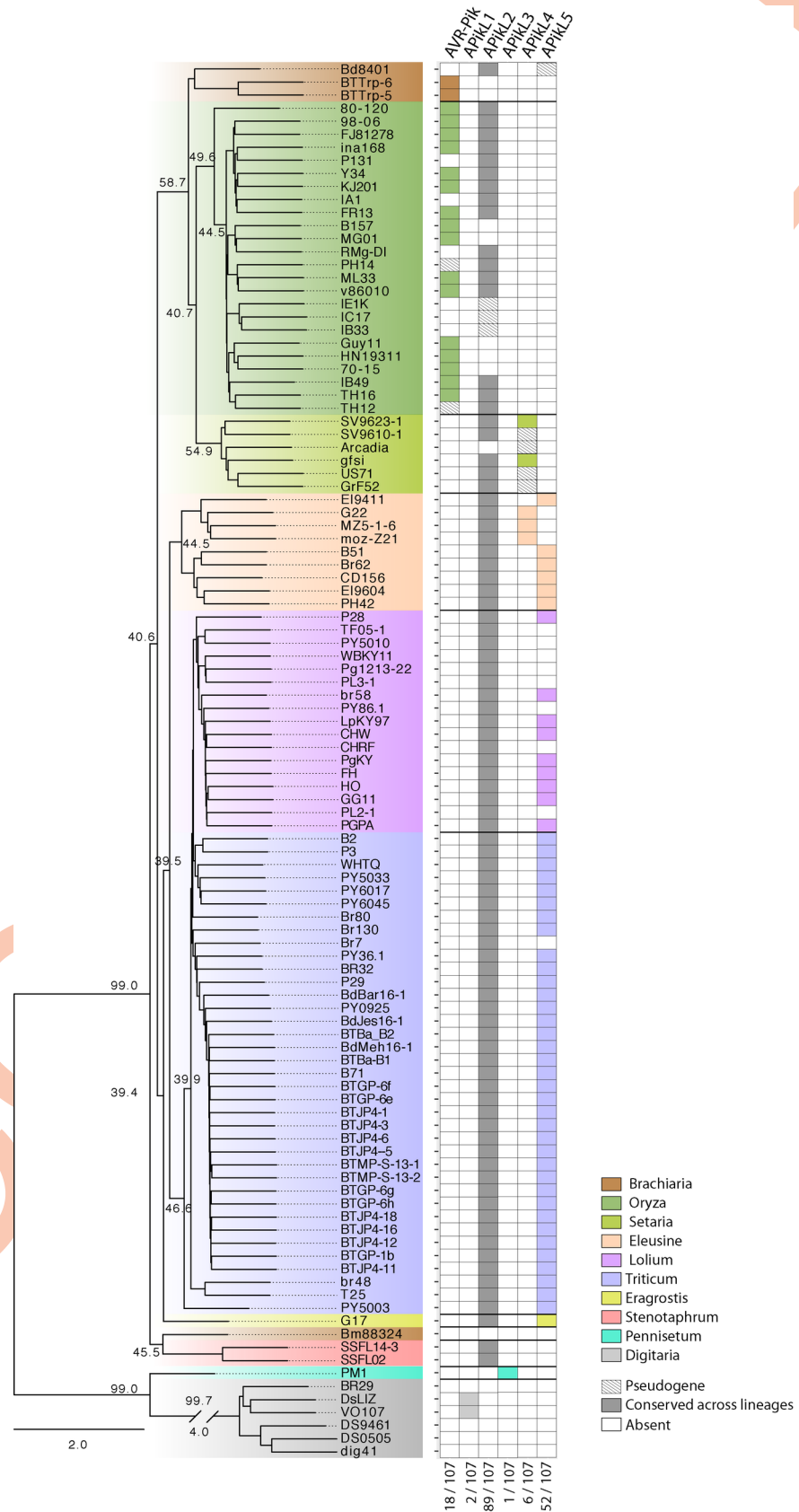


Fig 1. APikL2 is conserved across different host-specific lineages of *M. oryzae*. Presence/absence analysis of APikL family members shows conservation of APikL2 across all host-specific lineages of *M. oryzae*. All other family members show presence/absence polymorphisms that correlate with the species phylogeny and host-specificity. Left: ASTRAL multispecies coalescence tree derived from 1920 maximum-likelihood trees of all conserved single copy orthologs. Numbers at selected main branches represent the q1 local quartet tree support of individual genealogies. Right: Presence of APikL family members in each isolate. Colors indicate the host-specific lineages. The numbers at the bottom show the number of isolates that contain a certain effector.

<https://doi.org/10.1371/journal.ppat.1009957.g001>

polymorphisms in APikL effectors, we analysed allelic diversification within the APikL family. We used the full-length amino acid sequences of the 169 intact APikL family members to generate a maximum likelihood tree of the APikL family and map out allelic variants (Fig 2). We found one variant for APikL1 and APikL3, two variants for APikL4, four variants for APikL5, and six variants for APikL2 in addition to five known variants of AVR-Pik. The overall diversification of the whole APikL family resembled that of AVR-Pik which is known to be crucial for determining host compatibility driven by the co-evolution with the resistance locus Pik. Notably, we found that, although APikL2 is exceptionally conserved across all host-specific lineages of *M. oryzae*, the pattern of allelic diversification largely associated with particular lineages (Fig 2), similar to the pattern of presence/absence polymorphisms noted earlier for the rest of the APikL family.

The APikL2 gene locates to a conserved region on chromosome 3 in diverse genetic lineages of *M. oryzae*

We explored whether the genetic conservation of APikL2 is reflected in its genomic environment. To gain insights into the genomic location and dynamics surrounding the loci of APikL family members, we took advantage of several near chromosome quality assemblies, including eight isolates representative of five genetic *M. oryzae* lineages that contain the APikL members AVR-Pik, APikL2, APikL4, and APikL5. We first generated whole genome alignments of each assembly using isolate 70–15 (EnsemblFungi assembly MG08 [26]) as reference. We then identified homologous chromosomes or contigs in all assemblies by pairwise comparison to 70–15 and mapped the genomic location of APikL genes to each alignment (S1 Fig). All analysed isolates contain the APikL2 gene in a region approximately 1.5 MB from the start of chromosome 3 (S1 Fig), except the reference genome of 70–15 or the closely related isolate Guy11 that have lost APikL2.

The region around the APikL2 locus displays intra- and inter-chromosomal inversions as well as deletions that likely resulted in loss of APikL2 in 70–15 and Guy11 (S1 Fig). However, these rearrangements do not affect the overall location of the APikL2 locus on chromosome 3. Conversely, the genomic location of other APikL members is more variable. AVR-Pik localizes to the end of chromosome 2 in 70–15, to a contig with local similarities to chromosome 1 in FJ81278, to both ends of the mini-chromosome of FR13 [53], and to the end of chromosome 7 in Guy11 that likely resulted from an inter-chromosomal translocation between chromosome 2 and chromosome 7. Intriguingly, APikL4 and APikL5 also locate to the end of chromosome 7 in the isolates MZ5-1-6 and CD156 or B71, respectively, whereas APikL4 localizes to chromosome 2 in the *Setaria* infecting isolate US71. This analysis revealed that the genomic location of APikL2 across the examined isolates is conserved, whereas other APikL genes locate to variable regions in the genome.

To gain a more detailed view of the conservation and synteny around APikL genomic loci, we extracted the homologous contigs containing APikL2 or other APikL family members and performed pairwise alignments (Fig 3). This analysis revealed that frequent inversions are characteristic of the APikL2 locus, whereas the majority of chromosome 3 is highly syntenic

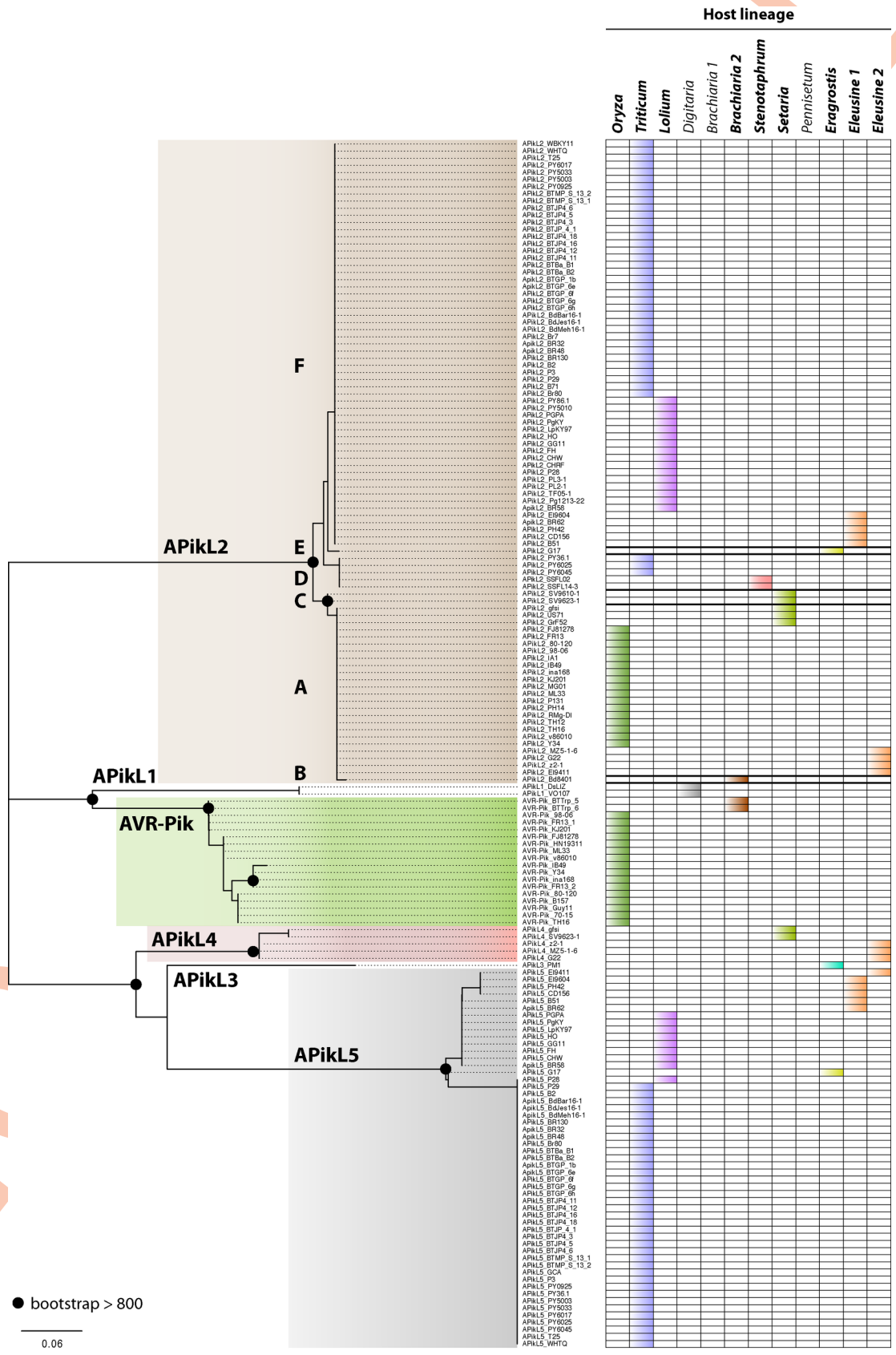


Fig 2. APiKL2 sequence diversification is associated with host-specific lineages. Left: APiKL family maximum likelihood tree based on amino acid sequences of all APiKL proteins from 107 *M. oryzae* genomes. Colors indicate APiKL family members where allelic diversification was observed in our dataset. Black circles indicate a bootstrap support >800 of the major nodes of allelic variants. Right: Presence of APiKL family members in various host-specific lineages. Colors indicate host lineages as shown in Fig 1. Host-lineages in bold indicate presence of APiKL2.

<https://doi.org/10.1371/journal.ppat.1009957.g002>

(Fig 3A). A similar pattern was present at the end of chromosome 7 where insertion and deletion events and an inversion surround the genomic region of APiKL4, APiKL5, and AVR-Pik (Fig 3C). We then extracted a 100 kb region surrounding the APiKL loci and analysed the conservation and colinearity of all genes within the regions. We mapped the B71 gene models (according to [28]) to the extracted regions and visualized gene conservation. This analysis showed that APiKL2 locates to a conserved region that contains 32 predicted genes from the B71 assembly. The majority of these genes are conserved and colinear between all analysed isolates and likely represent a secondary metabolite cluster (S2 Table). In isolates US71 (*Setaria* infecting lineage) and MZ5-1-6 (*Eleusine* infecting lineage), a ~30 kb insertion is present between APiKL2 and the neighbouring gene BSY92_00422 (Fig 3B), but this insertion didn't disrupt the gene co-linearity. Conversely, APiKL5 resides within a gene poor region in B71 containing 16 genes, mainly of unknown function. Eight and nine of these genes are conserved between B71 and CD156 or MZ5-1-6, respectively. Even though, AVR-Pik also locates to the end of chromosome 7 in Guy11, none of the B71 genes from the APiKL4 or APiKL5 loci is present in the 100 kb region surrounding the effector gene (Fig 3D). In summary, these analyses showed that APiKL2 resides in a conserved region on chromosome 3 across diverse host-specific isolates, whereas the genomic environment surrounding other APiKL family members is markedly more variable.

APiKL2 displays patterns of positive selection

Effector diversification is often driven by co-evolution of the pathogen with its host, which imposes strong selection pressure on the pathogen. To gain insights into the evolution and putative host-selective pressure acting upon the APiKL family we generated multiple sequence alignments, reconstructed the APiKL phylogeny and inferred the selective pressures acting on APiKL family members. We applied branch models and pairwise comparison implemented in the package PAML v4.9 [54,55]. These detected signatures of positive selection in the branches leading to APiKL2 (ω (dN/dS) = 1.27/1.47; free-ratio model/multiple-ratio model), APiKL1 (ω = 1.75/1.43), and APiKL5 (ω = 1.19/1.05; Fig 4A) consistent with high dN/dS (ratio of non-synonymous and synonymous substitutions) in pairwise comparisons (Fig 4B).

Importantly, as previously observed for pathogen effector genes under extreme levels of selection [56–61], several terminal branches showed amino acid substitutions in the absence of synonymous changes (dS = 0; Fig 4B and 4C). The underlying reason for these dN/dS patterns is thought to be the extreme levels of selection imposed by the host and associated population bottlenecks (as discussed by [14]).

Differential interaction of APiKL2 variants with the HMA domain protein sHMA94

Given the APiKL2 sequence similarity to AVR-Pik, we hypothesized that it binds plant HMA domains. We therefore tested interaction of APiKL2A and AVR-PikD with a collection of 151 HMA domains from *Oryza*, *Setaria* and *Triticum* in a pairwise yeast two-hybrid screen. These included 42 and 96 HMA domains from *S. italica* and *O. sativa*, respectively, representative of the sHMA diversity in these species and 13 HMA domains from *T. aestivum* with high

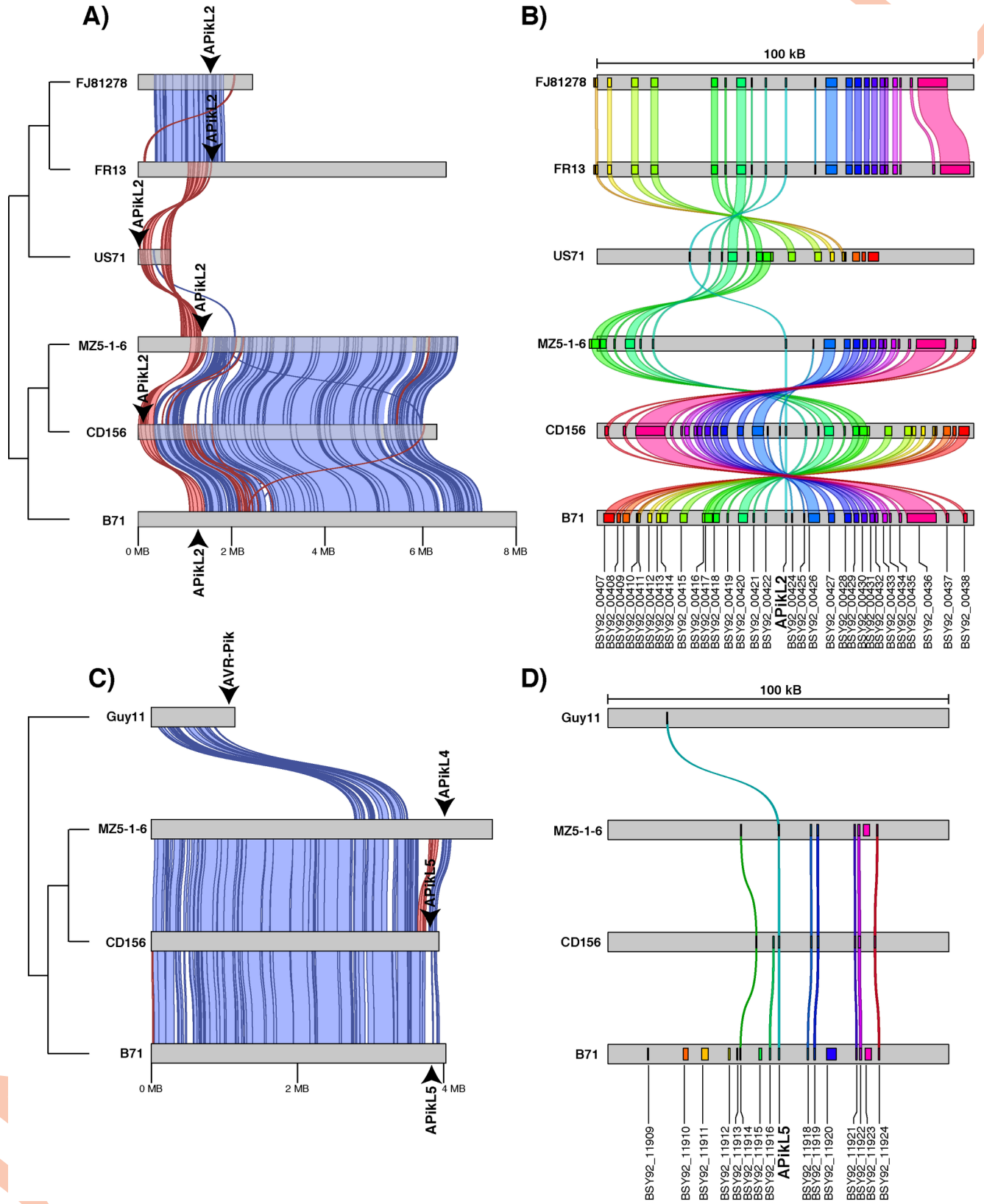


Fig 3. The APiKL2 locus is conserved across diverse genetic lineages of *M. oryzae*. Synteny analysis of APiKL loci in six isolates from five genetic *M. oryzae* lineages. A) Pairwise alignments of contigs containing the APiKL2 gene. Forward and reverse alignments are shown in blue and red, respectively.

Arrowheads indicate the position of the APiKL2 genes in the assemblies. **B)** Conservation of genes in a 100 kB region surrounding the APiKL2 locus. Colored links indicate homologous genes. **C)** Pairwise alignments of contigs containing diverse APiKL family members on chromosome 7. Forward and reverse alignments are shown in blue and red, respectively. Arrowheads indicate the position of APiKL family genes in the assemblies. Genetic relationship of the isolates is shown schematically on the left. **D)** Conservation of genes in a 100 kB region surrounding the loci of APiKL family members on chromosome 7. Colored links indicate homologous genes.

<https://doi.org/10.1371/journal.ppat.1009957.g003>

similarity to *O. sativa* clade A sHMAs [47]. This screen revealed 19 and 56 interactions for APiKL2A and AVR-PikD, respectively (S3 Table). APiKL2A interacted with 3, 5, and 11 HMA domains from *S. italica*, *T. aestivum*, and *O. sativa*, respectively. AVR-PikD interacted with 9, 7, and 40 HMA domains from *S. italica*, *T. aestivum*, and *O. sativa*, respectively. Overall, this screen revealed overlapping HMA-binding spectra of AVR-PikD and APiKL2A.

Next, we asked whether host-specific variants of APiKL2 differentially interact with HMA domains. APiKL2 is invariant in *Oryza*-infecting isolates but diverged into six variants across host-specific lineages of *M. oryzae* (Fig 2). The most divergent variant to APiKL2A is APiKL2F, present in *Oryza*- and *Triticum/Lolium*-infecting lineages, respectively (Figs 2 and 3). We therefore tested differential interaction of APiKL2A and APiKL2F with the HMA domains from *S. italica* and *T. aestivum* in a pairwise yeast two-hybrid assay. This screen revealed that APiKL2A and APiKL2F differentially interact with one target HMA-domain from the host plant *S. italica*, Si944233294 (EnsemblPlants accession: SETIT_012363mg), designated from here on as sHMA94 (Fig 5). In yeast two-hybrid assays, sHMA94 showed strong interaction with APiKL2F but no detectable interaction with APiKL2A, which might reflect functional specialization between the two effector variants. Three additional HMA domains from *S. italica* (Fig 5) and five HMA domains from *T. aestivum* interacted with both APiKL2 variants (S3 Table).

Two polymorphic residues in APiKL2 define interaction specificity with sHMA94

Four amino acids are polymorphic between the APiKL2A and APiKL2F proteins. To analyse the effect of these polymorphisms on the differential interaction to sHMA94, we tested 12 out of 16 possible mutant combinations of APiKL2A and APiKL2F in pairwise yeast two-hybrid assays against sHMA94 (Fig 6). We also included the HMA domain of Si514802025 (hereafter called sHMA25; EnsemblPlants accession: SETIT_011333mg), which interacted with both APiKL2 variants in our previous screen, as a control to exclude that any observed effects are due to secondary effects, such as protein stability.

These assays revealed that two polymorphic sites at position 66 (Asp in APiKL2A vs. Asn in APiKL2F) and 111 (Leu in APiKL2A vs. Pro in APiKL2F) are crucial determinants for sHMA94 interaction specificity of APiKL2 variants (Fig 6). Mutants that carry the APiKL2F residues Asn-66 and Pro-111 (M4, M6 and M7 in Fig 6) interacted with sHMA94 independently of the other two polymorphisms. In addition, mutants that carry Asn-66 but retain the APiKL2A residue Pro-111 (M1-3 and M5 in Fig 6) displayed a weak interaction with sHMA94 further highlighting the importance of this residue.

Asn-66 facilitates APiKL2 binding to sHMA94

To validate the yeast two-hybrid assays and further investigate the affinity of APiKL2A and APiKL2F for sHMA94, we performed isothermal titration calorimetry (ITC) with recombinantly expressed proteins purified from *Escherichia coli* (Fig 6B). The change in heat (ΔQ) for the interaction between APiKL2A and sHMA94 was notably reduced compared to the interaction between APiKL2F and sHMA94, suggesting a significant change in the thermodynamics

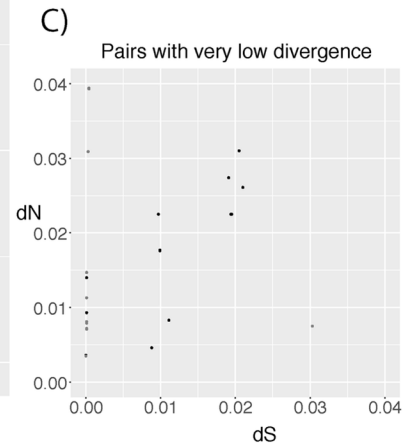
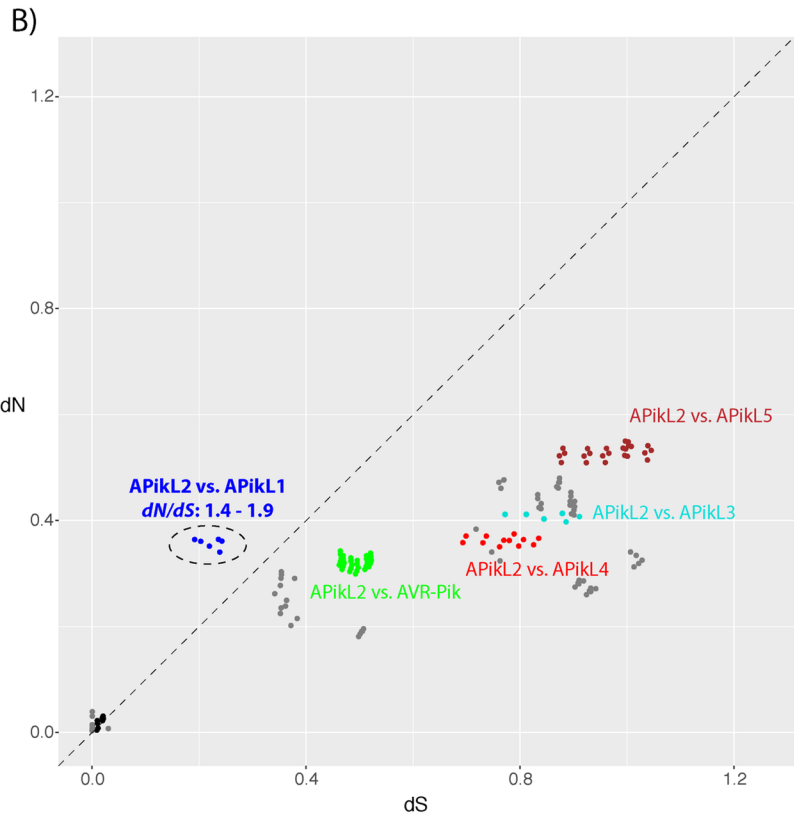
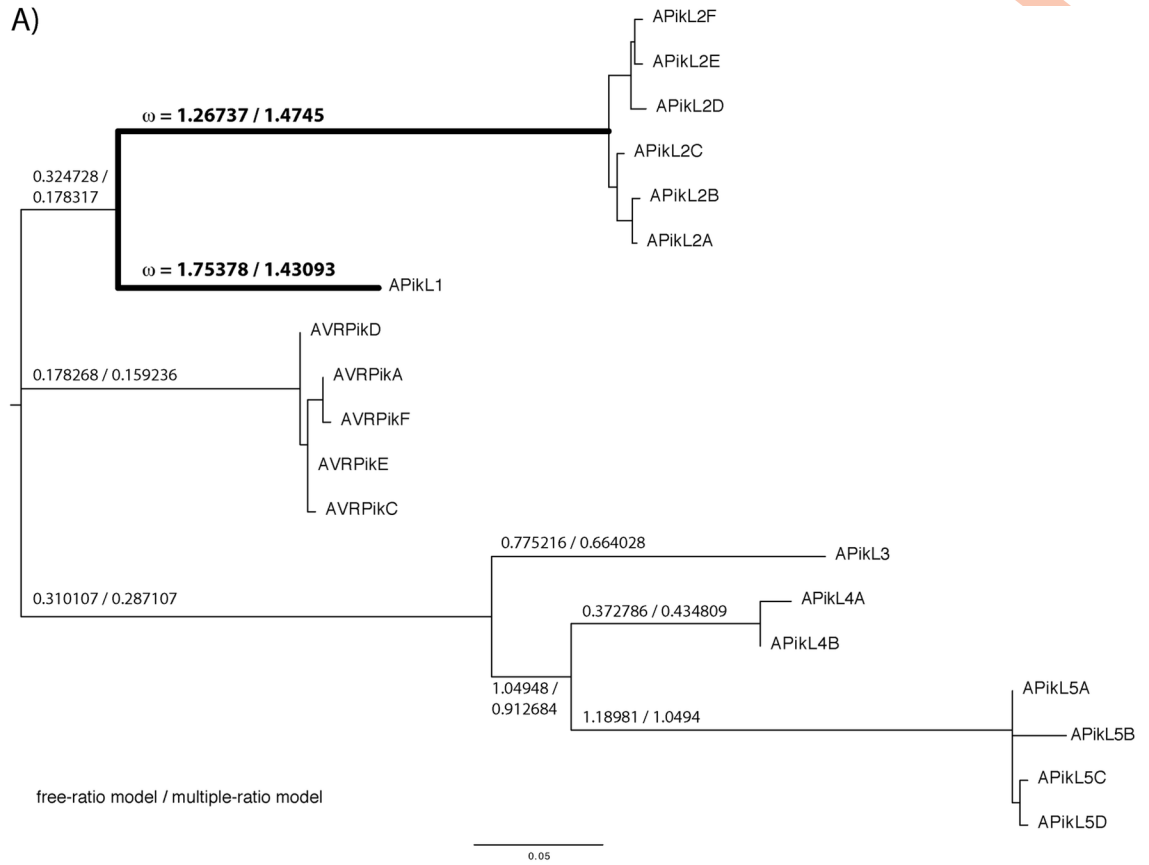


Fig 4. APiKL2 displays patterns of positive selection. **A)** Maximum-likelihood tree of all APiKL family members based on nucleotide sequences. Numbers on branches of the tree represent ω (dN/dS ratio of non-synonymous and synonymous substitutions) as calculated according to the free-ratio and multiple-ratio model implemented in PAML. Branches with high ω leading to APiKL2 and APiKL1 are highlighted in bold. **B)** Comparison of pairwise dN/dS ratios between APiKL family members. Colors indicate pairwise comparisons between APiKL2 and other APiKL family members. dN/dS ratios between other APiKL family members are shown in grey. Pairwise comparisons between APiKL2 variants are shown in black. The dotted line indicates $dN/dS = 1$. **C)** Enlarged view of pairwise comparisons in B with low divergence.

<https://doi.org/10.1371/journal.ppat.1009957.g004>

of the interaction. Consequently, we could only measure binding affinity of APiKL2F to sHMA94 with a calculated K_d of 145.8 nM. For APiKL2A we could not determine an accurate K_d , however, it is possible that it interacts weakly with sHMA94.

Next, we tested the influence of the key polymorphisms identified in the yeast-two hybrid mutant screen on in vitro sHMA94 binding. We first measured the effect of the Asp to Asn mutation at position 66 in APiKL2, as this is the only polymorphism between APiKL2A and APiKL2F that is predicted to locate to the HMA binding interface (based on [7,10,46]). Introduction of Asn-66 into APiKL2A was sufficient to facilitate sHMA94 binding, with a K_d of 384.6 nM similar to APiKL2F (Fig 6B). By contrast, the Asp-66 mutant reduced the affinity of APiKL2F for sHMA94, resulting in no measurable binding of sHMA94, similar to APiKL2A. Furthermore, the ΔQ of the interactions between the mutants and sHMA94 mirrored the changes in affinity, with the interactions between APiKL2A Asn-66 mutant and sHMA94 exhibiting a larger ΔQ , akin to wild-type APiKL2F, and APiKL2F Asp-66 mutant showing a significantly reduced ΔQ , similar to wild-type APiKL2A. While the yeast two-hybrid results also suggested an important role for Pro-111 (L111P in mutants M4, M6, M7) in sHMA94 interaction, we did not test this mutant in ITC as the single Asp-66 to Asn-66 mutation in APiKL2A

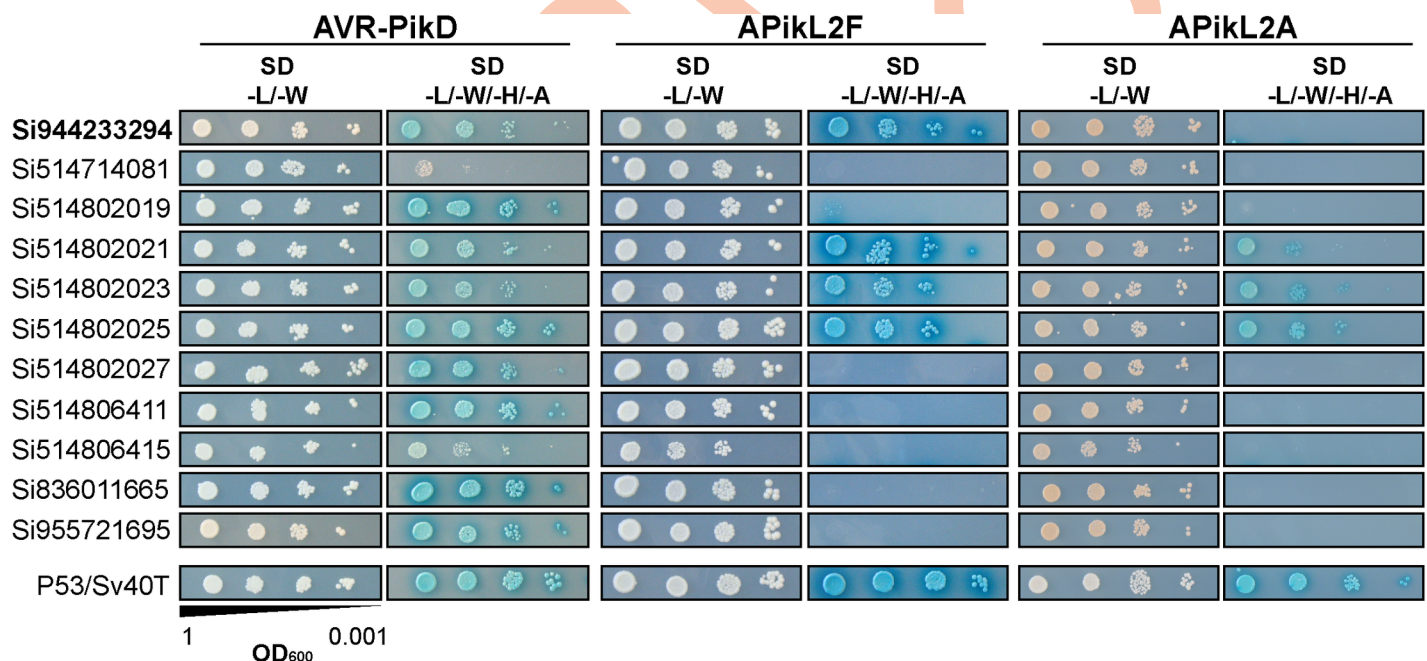


Fig 5. APiKL2 variants interact differentially with the HMA-domain of *Setaria italica* protein sHMA94. Subset of the pairwise yeast two-hybrid screen to identify interaction partners of AVR-Pik and the APiKL2 variants APiKL2A and APiKL2F from the *Oryza*- and *Triticum*-infecting lineages of *M. oryzae*. All positive interactions between *S. italica* HMA-domains with the APiKL family members AVR-Pik and APiKL2 are shown. Left: gene identifiers of HMA-domain containing proteins. SD -L/-W: Synthetic double dropout medium lacking the amino acids leucine and tryptophan; SD -L/-W/-H/-A: Synthetic quadruple dropout medium lacking the amino acids leucine, tryptophan, histidine, and adenine.

<https://doi.org/10.1371/journal.ppat.1009957.g005>

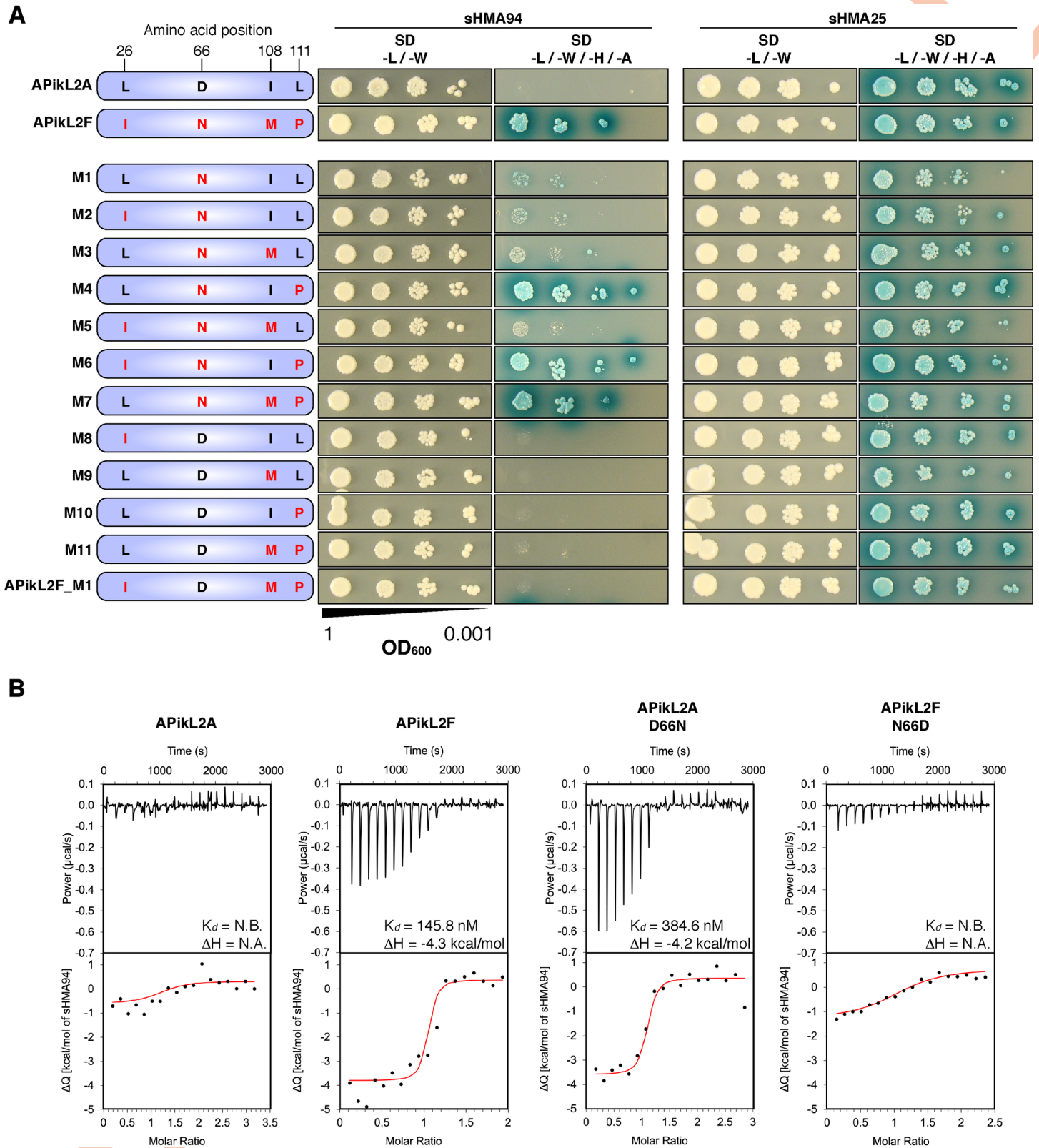


Fig 6. Two polymorphic sites determine binding specificity of APiKL2 variants to sHMA94. A) Pairwise yeast two-hybrid assays between sHMA94 and APiKL2A, APiKL2F, and 12 APiKL2 mutants derived from combinations of the four polymorphic residues. The two naturally occurring variants, APiKL2A and APiKL2F, were used as positive and negative controls. Mutations in each variant are shown schematically (left). sHMA25 was used as a positive control for both APiKL2 variants. SD -L/-W:

Synthetic double dropout medium lacking the amino acids leucine and tryptophan; SD -L/-W/-H/-A: Synthetic quadruple dropout medium lacking the amino acids leucine, tryptophan, histidine, and adenine. **B**) Asn-66 is essential for binding of APikL2 to sHMA94 *in vitro*. *In vitro* binding affinities for sHMA94 and APikL2 proteins were assessed by ITC. The top panels show heat differences upon injection of the effector into the cell containing sHMA94. The lower panel shows integrated heats of injection (•) and the best fit (red line) to a single site binding model calculated using AFFINImeter ITC analysis software. APikL2F, but not APikL2A, binds sHMA94. The APikL2F Asn-66 to Asp-66 polymorphism disrupts sHMA94 binding. Conversely, the APikL2A Asp-66 to Asn-66 polymorphism facilitates binding of sHMA94, with affinity similar to APikL2F.

<https://doi.org/10.1371/journal.ppat.1009957.g006>

restored sHMA94 binding to similar levels as APikL2F. From these data, we conclude that Asn-66 in APikL2 facilitates binding to sHMA94 compared to Asp-66.

Crystal structures reveal that Asn-66 stabilizes the binding interface between APikL2 and sHMA94

To better understand the molecular basis of interactions between APikL2 variants and sHMA proteins we sought to determine crystal structures of complexes. We crystallized complexes of APikL2F with sHMA94 and APikL2A with sHMA25 following co-expression in *E. coli* for comparative structural analysis. We were unable to crystallize a complex of APikL2A with sHMA94 for further study. X-ray diffraction data were collected at the Diamond Light Source (Oxford, UK) to 2.3 Å and 1.8 Å resolution, for APikL2F/sHMA94 and APikL2A/sHMA25, respectively. The structures were solved and refined as described in the Materials and Methods (S4 Table). In each case, the APikL2 variants and HMAs form 1:1 complexes, with both displaying the same global fold and interaction interfaces as previously described for the interaction of AVR-Pik with the Pik-1 integrated HMA domains [7,10] (Fig 7). One exception is the presence of an ordered N-terminal helix for APikL2, which is positioned away from the HMA-binding interface, but is disordered in AVR-Pik. The APikL2F/sHMA94 and APikL2A/sHMA25 complexes are very similar to each other (they superimpose with root-mean-square-deviation (RMSD) of 0.582 Å across the aligned C α backbones). Further, both complexes are very similar to the structure of Pikp-1 HMA/AVR-PikD (PBD ID: 6G10) and could be superimposed on this structure with RMSD's of 0.9 Å and 1.04 Å, respectively.

As we were unable to crystallize the complex of APikL2A with sHMA94, we compared the APikL2F/sHMA94 and APikL2A/sHMA25 structures here (sHMA25 interacted with both APikL2 variants in our yeast two-hybrid assays). The APikL2F/sHMA94 and APikL2A/sHMA25 complexes display the three binding interfaces previously reported for AVR-Pik/HMA complexes [7,10]. They share similar interface areas of ~964 Å and ~1011 Å (calculated by QtPISA [62]) contributed by 57 and 56 residues, respectively. They also form a similar number of hydrogen bonds, salt bridges, and maintain a similar predicted binding energy (S5 Table). The largest differences between the complexes are within interface 3, where APikL2 forms several interactions with sHMA25 that cannot be formed with sHMA94 due to differences in the sequences of the HMAs in this region (S2 Fig). The interface 2 region of sHMA94 and sHMA25, which forms contacts with the Asp-66/Asn-66 residues of APikL2, is highly conserved in both complexes.

However, we observed that the presence of Asn-66 in APikL2F results in an altered local hydrogen bonding network around Gln-44 in sHMA94, when compared to Asp-66 in the APikL2A/sHMA25 complex (Fig 8). As a result, the side chain of Gln-44 of sHMA94 is rotated relative to Gln-43 of sHMA25, allowing Gln-44 to form additional intramolecular hydrogen bonds with Thr-38 in sHMA94 (Fig 7B and 7D). In this orientation, the NE2 atom of Gln-44 interacts with the OG1 atom and backbone carbonyl of Thr-38. This hydrogen bonding arrangement is not adopted in the APikL2A/sHMA25 complex, as the Asp-66 in APikL2A forms a hydrogen bond with the NE2 atom of Gln-43 in sHMA25 (Figs 7A, 7C and S3), rotating the NE2 group away from Thr-37, and preventing formation of hydrogen bonds between

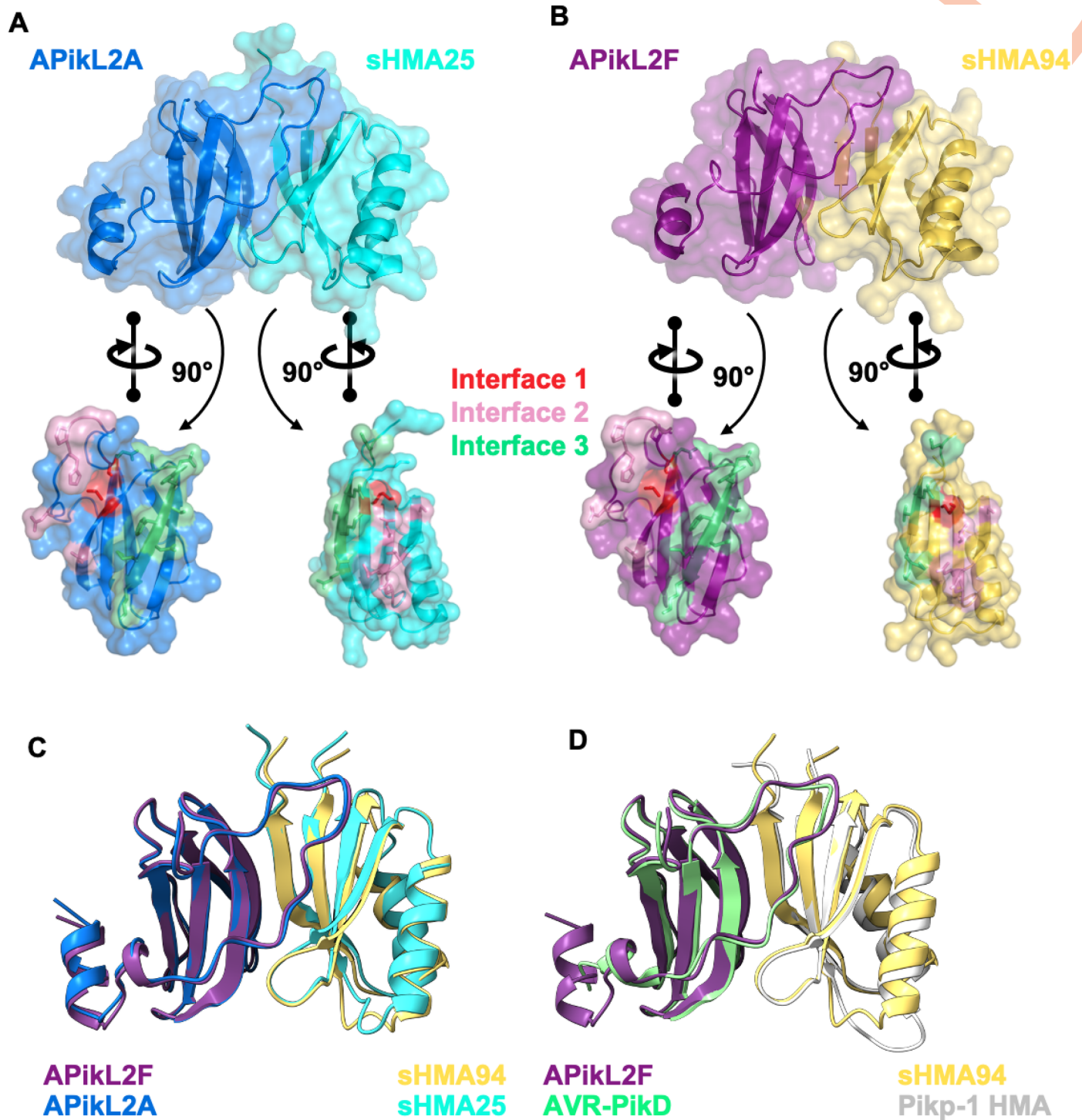


Fig 7. Crystal structures of APiK2 in complex with HMA domains reveals a conserved binding interface. A) Cartoon and surface representation of the APiK2A/sHMA25 complex, with APiK2A coloured blue and sHMA25 coloured cyan. B) Cartoon and surface representation of the APiK2F/sHMA94 complex, with APiK2F coloured purple and sHMA94 coloured yellow. The effector was separated from the complex using PyMol and rotated 90° relative to the complex to give an interior view of the interaction interface. The residues that form the interfaces one, two, and three (as defined by [7]), between the effector and HMA are coloured dark blue, pink, and magenta, respectively. C) Superimposition of the APiK2F/sHMA94 (purple and yellow) and APiK2A/sHMA25 (blue and cyan) complexes performed with PyMol. D) Structural alignment of APiK2F/sHMA94 (purple and yellow) with AVR-PiK/Pikp-1 HMA (green and grey; PDB ID: 6G10).

<https://doi.org/10.1371/journal.ppat.1009957.g007>

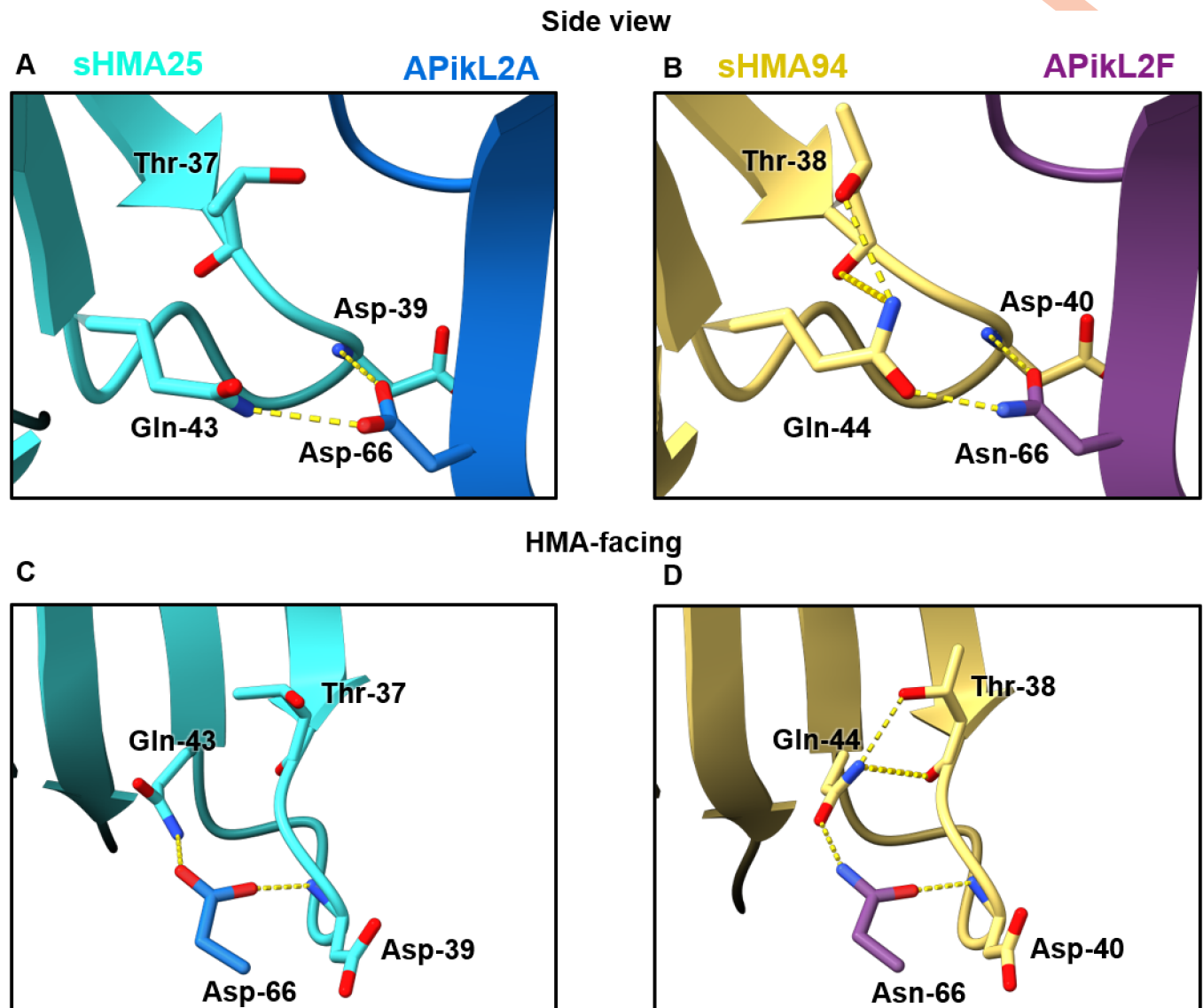


Fig 8. The aspartate/asparagine polymorphism at position 66 in APiKL2 results in an altered hydrogen bonding network at the sHMA interface. A, C) The carboxylic side chain of Asp-66 from APiKL2A forms two hydrogen bonds with sHMA25 through the NE2 atom of Gln-43 and the backbone nitrogen of Asp-39. **B, D)** The amide side chain of Asn-66 from APiKL2F forms two hydrogen bonds with sHMA94 through the OE1 atom of Gln-44 and the backbone amide of Asp-40. The interaction with Gln-44 via the OE1 atom in the AVR-Pik2F/ sHMA structure flips Gln-44 relative to the Gln-43 residue of the APiKL2A/sHMA25 structure. This reorientation of Gln-44 in sHMA94 allows the NE2 atom of Gln-44 to form intramolecular hydrogen bond with the OG1 atom and backbone carbonyl of Thr-38.

<https://doi.org/10.1371/journal.ppat.1009957.g008>

these two residues. The altered hydrogen bonding network around Gln-44 in the APiKL2F/ sHMA94 complex, compared to APiKL2A/sHMA25, appears to firmly lock Gln-44 in place compared to Gln-43 of sHMA25. This suggests that the interaction with Asn-66 of APiKL2F results in reduced flexibility of the sHMA94 loop in interface 2. Our data indicates that this hydrogen bonding network stabilizes the interaction between APiKL2F and sHMA94, providing an explanation for the binding specificity observed in the yeast two-hybrid and *in vitro* binding affinity assays. However, it is likely that the three interfaces differentially impact binding to divergent HMA-domains. Consistent with this, both APiKL2 variants can bind to

sHMA25 where hydrogen bonds in interface 1 and 3 must have a stronger impact on complex formation compared to sHMA94.

Asn-66 is a recently derived polymorphism in APiKL2 effectors

Our results show that APiKL2 and AVR-Pik share a common fold and HMA domain binding interfaces. This suggests that all APiKL family members share a common fold and that specialization to host target HMAs is facilitated by adaptive mutations in common interfaces. To investigate this, we first mapped the evolutionary history of the polymorphisms that define the effector diversity in natural populations of *M. oryzae* and reconstructed the ancestral states of APiKL family members (Fig 9A). This revealed that successive mutations have led to host-specific alleles for each of the APiKL effectors. Consistent with tests for positive selection, most of the changes that contribute to allelic diversity in terminal branches are not associated with synonymous changes, similar to AVR-Pik and other AVR-effectors under strong positive selection [56–61]. The ancestral reconstructions further revealed the stepwise order of emergence of the crucial polymorphisms that determine binding specificity in both APiKL2 and AVR-Pik (Fig 9A). The ancestral state at position 66 is Asp, which only mutated to Asn-66 in the branch leading to APiKL2F, APiKL2D and APiKL2E (Fig 9A). In contrast, Pro-111 is ancestral to Leu-111, which only emerged in the APiKL2A and APiKL2B branch (Fig 9A). These results indicate that APiKL2 binding to sHMA94-like HMA domains is a recently derived trait that evolved through the Asp to Asn polymorphism at position 66. We conclude that the most plausible evolutionary scenario is that Asn-66 enabled the APiKL2 effector to expand its binding spectrum to sHMA94-like target proteins.

Diversifying residues are centred around the two HMA-binding interfaces across the APiKL effector family

The sequence diversity and crystal structures of the APiKL family provide a rare opportunity to interpret information from protein structures in an evolutionary context. To determine the spatial distribution of polymorphic residues, we mapped the conservation score for each amino acid (S4 Fig; inferred by ConSurf [63]) and residues contributing to allelic diversification (Fig 9B and 9C), onto the crystal structures of AVR-Pik and APiKL2, and homology models of APiKL4 and APiKL5. We noted that highly divergent residues (S4 Fig) primarily locate to the N-terminal region and disordered loop region of the effectors, whereas highly conserved residues tend to locate in the core structural β -sheets that contribute to the MAX-fold (S4 Fig), consistent with their role in maintaining structural integrity. Further mapping of non-synonymous mutations that contribute to allelic diversification in terminal branches of the APiKL family revealed two major regions that may modulate interactions with host proteins (Figs 9C and S5).

These regions are spatially consistent with two of the three HMA binding interfaces identified in AVR-Pik/HMA complexes [7,10,43], and APiKL2/HMA complexes detailed in this study. They locate to the N-terminal loop that forms interface 2 and to a surface-exposed patch formed by β -sheets 2 and 3 that corresponds to interface 1. In summary, these analyses indicate that two structurally distinct regions that map to HMA binding interfaces may contribute to adaptive evolution of the APiKL family.

Discussion

Accelerated gene evolution is a hallmark of pathogen adaptation and is often associated with host-range expansions and host-jumps [1,14–21]. However, despite extensive knowledge about the evolutionary dynamics of pathogen effectors, the molecular and biochemical basis of

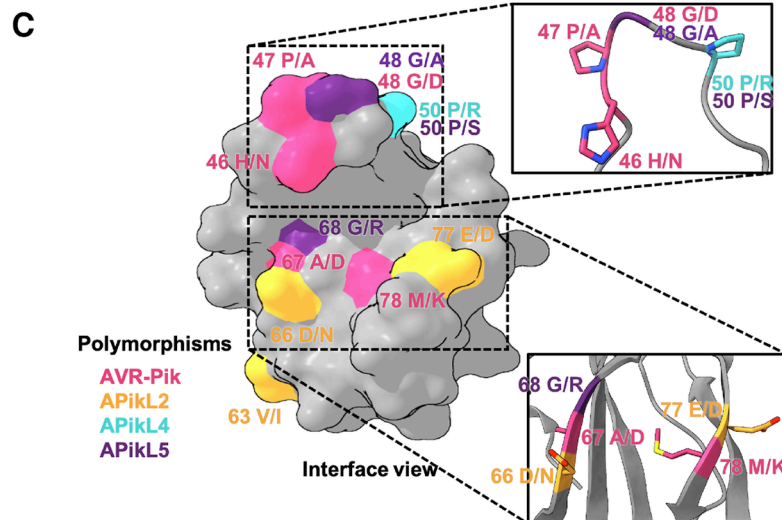
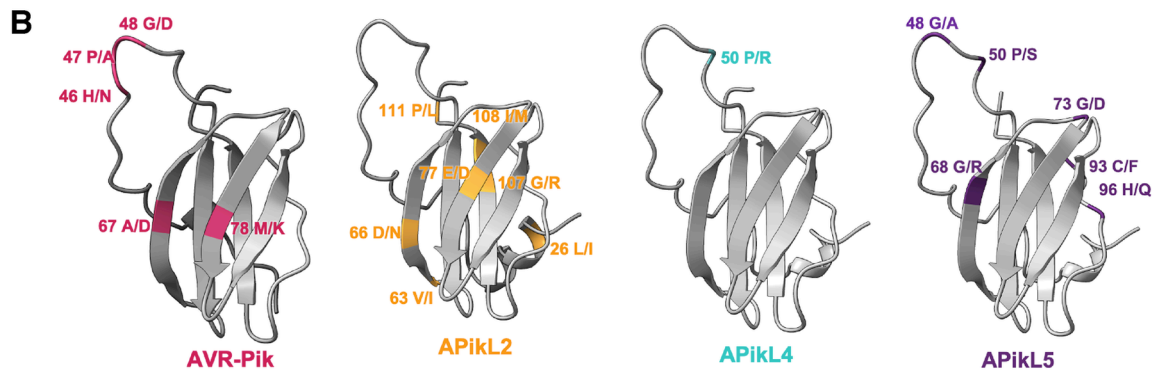
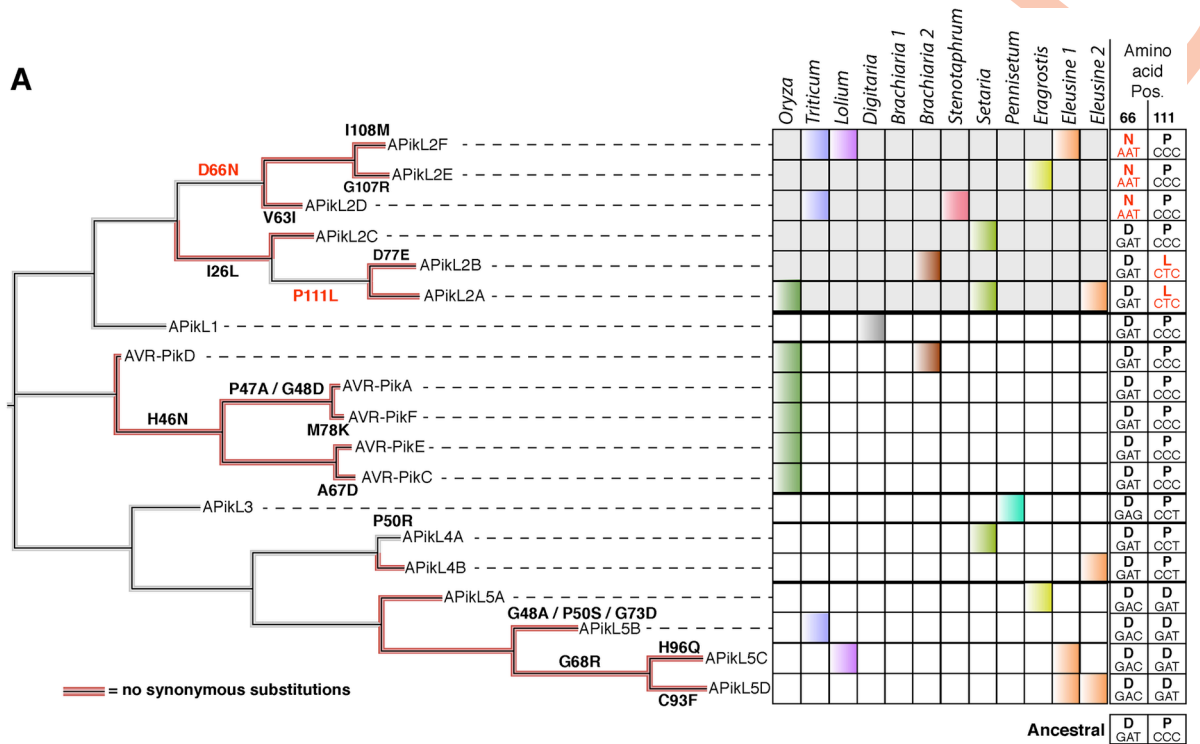


Fig 9. Adaptive mutations shape the binding interfaces of APiKL family effectors. A) Asn-66 is a recently derived polymorphism in APiKL2 effectors. Left: Reconstruction of the evolutionary history of the APiKL family. Multiple non-synonymous mutations contribute to allelic diversification, largely in absence of synonymous mutations (red branches). Right: APiKL effector presence/absence across multiple host-specific lineages. Colors correspond to host lineages as in Fig 1. Amino acids and codons of positions 66 and 111 across the APiKL family are shown on the right. Acquired mutations are highlighted in red. The ancestral states are shown at the bottom. B) Mapping of non-synonymous changes that contribute to allelic diversification to the structures of APiKL family members AVR-Pik, APiKL2, APiKL4, and APiKL5. Divergent amino acids of each effector are highlighted. C) Allelic diversification reveals two major hotspots across the APiKL family. Surface representation of divergent sites across the APiKL family combining all residues shown in B. Most polymorphic residues reside in the HMA-binding interfaces 1 (bottom) and 2 (top).

<https://doi.org/10.1371/journal.ppat.1009957.g009>

how they adapt to new host-targets is still largely unknown. Here, we investigated a family of putative effector proteins with sequence similarity to the well-characterized rice blast fungus effector AVR-Pik. We found that all six members of this effector family show presence/absence polymorphisms and allelic diversification in various lineages of the multihost blast fungus. We demonstrated that allelic variants of the exceptionally conserved effector, APiKL2, differentially bind the host HMA domain protein sHMA94. By combining crystal structure information of two divergent APiKL2 variants in complex with HMA-domains and ancestral reconstruction of the APiKL family, we showed that three APiKL2 variants, APiKL2D/E/F, acquired a derived mutation—Asn-66—that maps to the HMA binding interface and expands the effectors binding spectrum to the new target, sHMA94. We further showed that polymorphic residues under positive selection in the APiKL family tend to locate at two major HMA-binding interfaces, suggesting that adaptation and coevolution with plant HMA domains may drive the evolution of this effector family. Taken together, our results provide a detailed molecular evolution and structural biology framework for diversification and adaptation of effectors across host-specific lineages of the blast fungus *M. oryzae*.

Effector proteins often interact with multiple members of host-target protein families [47,64–66]. In addition, multiple pathogen effectors can target the same host protein family [35,36,67], possibly to ensure effective host infection or as an evolutionary “bet-hedging” strategy to counteract detection by the plant immune system [66]. APiKL2 shares the MAX-effector fold and similar HMA-binding interfaces with the well characterized effector AVR-Pik. In addition, homology modelling of APiKL family members shows that most APiKL effectors probably carry adaptive mutations within these binding interfaces (Fig 9). Given this, and the similar degree of sequence conservation across the APiKL family (with the exception of APiKL5 which contains a 19 amino acid C-terminal extension), we hypothesize that all APiKL family members bind to multiple sHMA proteins via a conserved mode of action. Such a high degree of functional redundancy, along with the HMA-binding previously demonstrated for the MAX-effectors AVR-Pik, AVR1-CO39 and AVR-Pia [35], further indicates that targeting HMA proteins is an important virulence activity during host infection by the blast fungus. Future genetic analyses (gene knock-out and complementation) would need to take into account possible functional redundancy among the effectors as well as the diversity of sHMA proteins in grass hosts. In addition, we noted variation in gene expression of APiKL genes in the limited public datasets that are available. In the future, comprehensive gene expression analyses of the APiKL genes across multiple lineage/host combinations may reveal another layer of diversification in these effectors.

Our findings on the APiKL family are consistent with the view that these effectors collectively target sHMA proteins of their multiple grass hosts with varying degrees of redundancy and specificity. We propose that diversification of the APiKL family might be adaptive and driven by co-evolution with HMA containing proteins. Although the canonical member of this family, AVR-Pik, is restricted to the *Oryza*-infecting lineage of *M. oryzae*, the other five members of the family, APiKL1 to APiKL5, are widely distributed across other host-specific

lineages of *M. oryzae* (Figs 1 and 2). AVR-Pik, APikL2, and possibly the other APikL effectors, are structurally similar and bind HMA domains through common interfaces (Fig 9). APikL2 and AVR-Pik have overlapping spectra of HMA interactors indicating partially redundant interactions with this host target family. Interestingly, most *M. oryzae* isolates contain two members of the APikL family, with APikL2 present in isolates from all genetic lineages whereas the other APikL effectors being specific to certain host-specialized lineages. Presence of two partially redundant effector proteins might expand the spectra of sHMA proteins that are targeted during infection. In addition, the patterns of presence/absence polymorphisms observed for the APikL family across lineages of *M. oryzae* might reflect recurrent gene losses to evade plant immunity whenever one of the effectors is detected. This work sets the stage for future functional studies on the precise role of APikL effectors in host-specificity and incipient speciation of *M. oryzae*.

The six APikL2 variants grouped into two clades APikL2A-C and APikL2D-F. Particular alleles tend to be associated with a given host-specific lineage. For example, the *Oryza*-, *Setaria*-, and *Brachiaria*-infecting lineages only encode APikL2A-C and lack the other APikL2 variants, which carry the D66N polymorphism (Fig 9). Interestingly, *Eleusine*-infecting lineages are the only ones to carry both APikL2A and APikL2F variants and thus both, Asp-66 and Asn-66 (Fig 10). In addition, the *Eleusine*-infecting lineages have the most diverse repertoire of APikL effectors, carrying three APikL homologs, APikL2, APikL4, and APikL5 (Figs 1, 2, and 3). This indicates that APikL2 diversification and APikL4/5 gene gain/loss may have occurred in ancestral *Eleusine*-infecting populations prior to the emergence of the distinct genetic lineages and sorting of the various APikL genes. However, we cannot exclude recurrent host jumps between *Eleusine* and other host species as previously reported [68].

This model is further complicated by the presence of lineages of *M. oryzae* that are associated with crop plants but retain the capacity to infect wild grass hosts. For example, isolates from wild populations of *E. indica* may infect the cultivated relative *E. coracana* and vice versa (Fig 1 and [23]). Specifically, isolate PH42 which was collected from *E. coracana*, groups with *E. indica* infecting isolates in our analysis (Fig 1 and S1 Table). Conversely, isolate EI9604, collected from *E. indica* groups with isolates infecting *E. coracana* suggesting cross-infectivity between wild and cultivated relatives (Fig 1 and S1 Table). Cross-infectivity of different hosts has also been shown for multiple lineages under laboratory conditions [69]. It is possible that shuttling between wild and cultivated plants, as in between *E. indica* and *E. coracana*, could have facilitated effector diversification and such cases can serve as springboards for host range expansions and host-jumps of the blast fungus [68]. This calls for increased surveillance of wild grass hosts, especially in regions affected by emerging blast disease [70–73], in order to understand the role they play in host-driven adaptation of *M. oryzae*.

We found that a single amino acid—Asp-66 to Asn-66—in the APikL2D-F variants leads to an expansion of the HMA-binding spectrum and that the derived Asn-66 polymorphism is likely an adaptive feature of the APikL2 variants of isolates from *Triticum*, *Lolium*, *Stenotaphrum*, *Eragrostis*, and *Eleusine*. What are the selective forces that may have driven the emergence of the Asp-66 to Asn-66 polymorphism? It is important to note that although Asn-66 only emerged in some APikL2 variants within the APikL effector family (Fig 9), it predates the split between *M. oryzae* lineages infecting *Triticum*, *Lolium*, *Stenotaphrum*, *Eragrostis*, and *Eleusine* (Fig 10). Therefore, it is difficult to ascertain which of these or any other ancestral host species may have exerted the selection pressure that led to the fixation of Asn-66. In addition, sHMA94, the newly gained host target of APikL2 variants carrying Asn-66 is from *Setaria italica*, and *Setaria*-infecting isolates do not carry Asn-66 but rather the ancestral Asp-66. Nonetheless, we propose that Asn-66 enabled APikL2 to expand its binding spectrum to new sHMA94-like proteins that probably occur in one or several of the *Triticum*, *Lolium*,

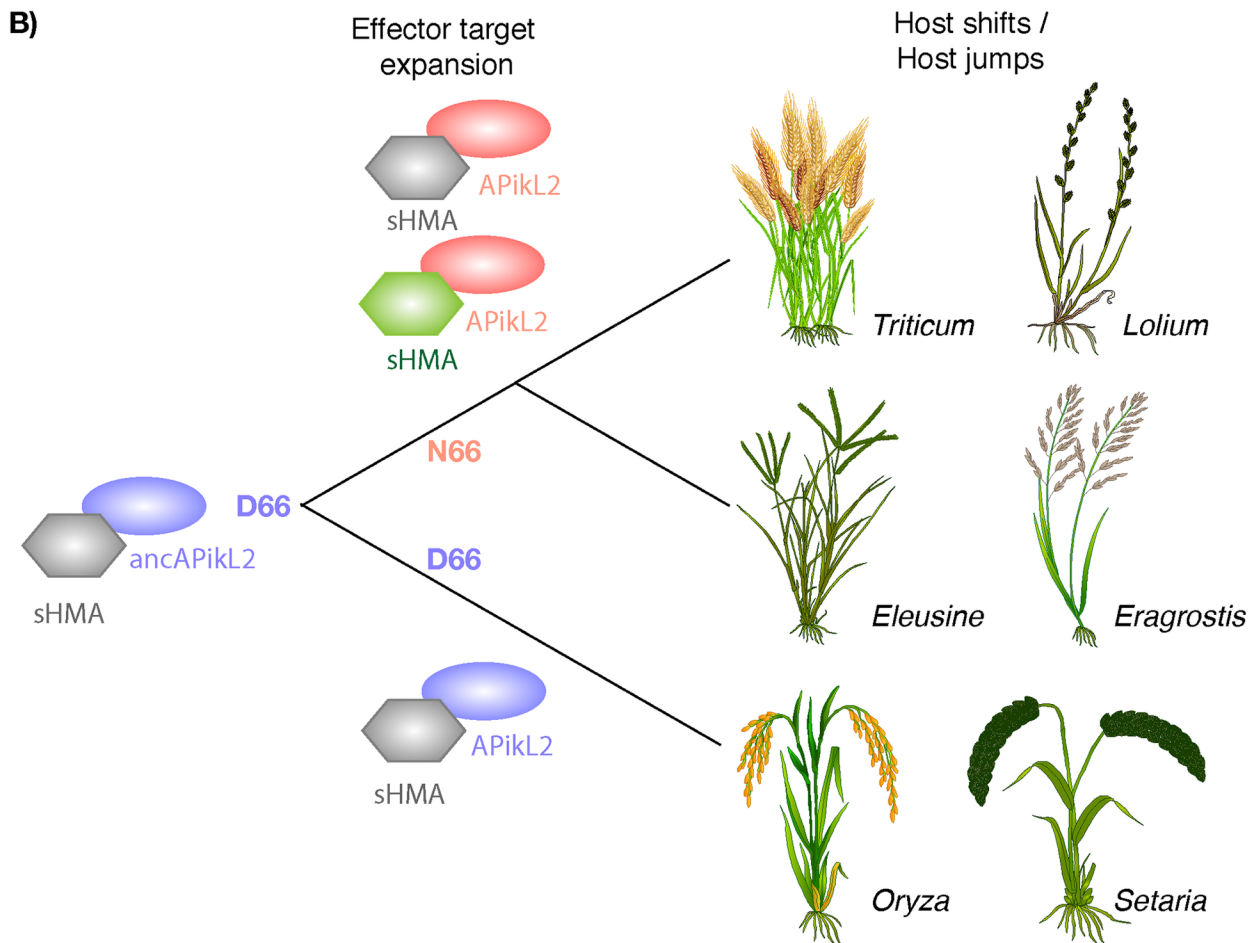
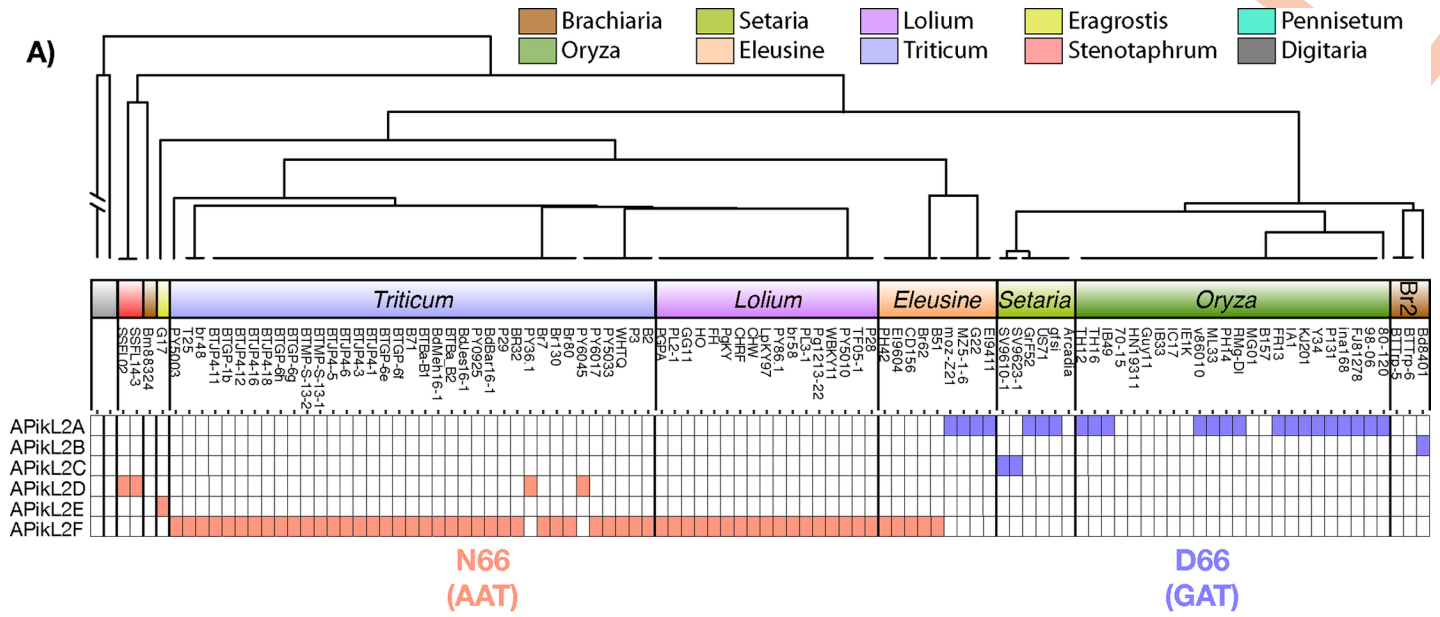


Fig 10. The Asp-66-Asn polymorphism predates the emergence of host-adapted lineages of *M. oryzae*. A) Distribution of the critical Asp-66 to Asn-66 mutation in *M. oryzae* populations. Schematic illustration of host-specific lineages of *M. oryzae* and annotation of the Asp-66/Asn-66 polymorphism. Blue: Asp-66; Red: Asn-66. The

host genera of the different *M. oryzae* lineages are shown. B) Model illustrating the emergence of the Asn-66 polymorphism prior to the differentiation of host specialized lineages. Blue: APikL2 variants carrying Asp-66; Red: APikL2 variants carrying Asn-66; Grey: sHMA target protein interacting with both APikL2 variants; Green: New sHMA target protein due to Asn-66 polymorphism.

<https://doi.org/10.1371/journal.ppat.1009957.g010>

Stenotaphrum, *Eragrostis*, and *Eleusine* hosts. We also conclude that it is unlikely that the Asp-66 to Asn-66 polymorphism resulted in evasion of detection by an immune receptor (*R* gene) given that it is difficult to envision how gain of binding to a new host target would evade immunity.

When we examined the structures of the APikL2/sHMA complexes we were surprised at the subtlety of the changes induced by the Asp-66 to Asn-66 polymorphism given the clear difference in sHMA94 binding we observed via yeast-2 hybrid and ITC. Comparison of the Asp-66 to Asn-66 polymorphism in the APikL2A and F alleles to the polymorphisms found in AVR-Pik alleles further highlights the discrete nature of the APikL2 polymorphism. AVR-Pik alleles that are not recognised by the Pik NLRs have reduced hydrogen bonding interactions between the effector and HMA domain as a result of their polymorphisms, which correlate with reduced affinity and loss of recognition by the receptor [7]. In contrast to this, net sum calculations of hydrogen bonds between structures of APikL2A and F bound to sHMA proteins do not explain why APikL2F has a greater affinity for sHMA94 than APikL2A, as both complexes maintain a similar number of contacts. ITC showed that APikL2A has a lower affinity for sHMA94, and the heat exchange (ΔQ) is decreased, implying significant changes in the enthalpy of the system. In correlation with this, we observe the formation of additional intramolecular hydrogen bonds around Gln-44 of sHMA94 in the presence of Asn-66 found in APikL2F, possibly reducing the mobility of Gln-44 and rigidifying this loop region of the interface. We hypothesise this rigidification of sHMA94, a result of the intramolecular hydrogen bonds in the presence of Asn-66, is intrinsic to the increased affinity of APikL2F. The hydrogen bonding network in this region may lock sHMA94 in a more favourable conformation for binding, leading to the increased affinity of the complex [74,75].

The subtlety of the molecular mechanisms underlying the APikL2F/sHMA94 interaction highlights the importance of experimental structure validation. Early attempts to predict the interactions between APikL2 and sHMA94 using homology modelling with AVR-Pik and the Pkp-1 HMA domain as templates gave little meaningful insight into the structural basis for the differential binding of APikL2A and APikL2F to sHMA94. Structural modelling of proteins to infer the impact of single residues on interactions in a complex is ultimately challenging, as side chain predictions are frequently made using backbone-dependent rotamer libraries [76,77], and do not sufficiently account for the interaction environment surrounding the residue at an interface. While protein complex modelling programs do exist, these also rely on the target maintaining significant sequence similarity with the template structure [78,79]. This presents significant hurdles for rapidly diversifying fungal effectors. While structural modelling can be used to identify putative functions and assist with categorization of sequence diverse effectors [80], characterizing the molecular details of effector-target interactions remains challenging with these approaches.

Despite major advances in understanding how pathogen effectors evade plant immunity, we still know little about how effectors acquire new host targets, especially as the pathogen shifts from one host to another [1,14–21]. Our study provides a molecular evolution and structural biology framework to study effector adaptation in a multihost pathogen. We conclude that the APikL family has diversified through a series of polymorphisms at common effector-target binding interfaces (Fig 8B and 8C). This diversification was associated with distinct patterns of genomic location with the conserved APikL2 effector residing in a highly co-linear

gene-rich region whereas the other effector genes occurring at different genomic loci in genetically distinct *M. oryzae* isolates (Fig 3). This work also illustrates the value of ancestral sequence reconstruction in yielding a more plausible model of plant-pathogen coevolution [1,43,81,82] and the unique perspectives gained from evolutionary molecular plant-microbe interactions (EvoMPMI) approaches [4,83]. By establishing the directionality of evolution and that D66 (GAT codon) is ancestral to N66 (AAT codon) (Figs 9A and 10), we can propose that the D66N polymorphism is more likely to reflect expanded host-target binding rather than evasion of host immunity. This enabled us to better understand the selection forces and historical events that have shaped the complex evolution of the APiKL effector family.

Material and methods

Inference of a coalescence species tree by ASTRAL

To infer the species tree, we first extracted all conserved single copy orthologs from 108 genomes, representative of ten host-specific lineages, using BUSCO (version 3) with the *sordariomycetes_odb9* database. Isolate PY6025 was excluded from further analysis due to low BUSCO completeness. We then used all 1920 universally conserved single copy orthologs to generate individual genealogies by RAXML (version 8.1.11) using the GTR-GAMMA model with 1000 bootstrap replications. We then used each individual best tree as input for ASTRAL (version 5.7.3). Local branch support was evaluated by local quartet tree support “q1” using the -t option in ASTRAL that reports the proportion of local quartets in the set of individual genealogies that agree with the topology of the species tree.

Identification of APiKL family effectors and potential HMA-target domains

We used an iterative TBLASTN approach to identify APiKL family members using 107 genomes of *M. oryzae*. In the first round, we used AVR-PikD as a query and selected all hits with 100% query coverage. This led to the identification of APiKL1-4. We then repeated the search using all APiKL family members as a query. Using APiKL4 as query led to the identification of the more divergent effector APiKL5. This approach led to the identification of 179 candidate effectors that included 10 pseudogenes. APiKL family members were classified using an amino acid percent identity cut-off of >90% for variants of the same effector.

HMA target domains were identified by TBLASTN using the integrated HMA-domain of the rice sensor NLR Pikp-1 as a query. Details about the target domains, sequences and plasmids used are available in [84]. All plasmids are available through Addgene (plasmid accession numbers 168306–168456).

Generation of the APiKL family tree

To infer the APiKL family tree we removed all pseudogene sequences from the analysis. We then created a multiple sequence alignment of 169 amino acid sequences using MUSCLE [85]. We then generated a maximum-likelihood tree with 1000 bootstrap replications in MEGA7 [86].

Protein-protein interaction

Yeast 2-hybrid interaction screen. We used the Matchmaker Gold Yeast two-hybrid system (Takara Bio USA) to detect protein-protein interactions between APiKL effectors and putative HMA-target domains following the manufacturer’s instructions. All effector sequences were codon optimized and synthesized for expression in *E. coli* and cloned into pGADT7 plasmids.

All HMA-domains were codon optimized and synthesized for *E. coli* expression and cloned into pGBKT7 plasmids. Each individual effector-HMA combination was then co-transformed into competent *Saccharomyces cerevisiae* Y2H Gold cells (Takara Bio USA) and selected on SD^{-Leu-Trp} (synthetic dropout) medium. Single colonies were selected and grown overnight at 28°C and 200 rpm to an OD₆₀₀ of 1–2 in liquid SD^{-Leu-Trp} medium. Cultures were then used to prepare serial dilutions of OD₆₀₀ 1, 1⁻¹, 1⁻², 1⁻³. Of each dilution 4 µl were replica plated on SD^{-Leu-Trp} and SD^{-Leu-Trp-Ade-His} containing the chromogenic substrate X-α-Gal. Pictures were taken after 72–96 h incubation at 28°C. Yeast 2-hybrid screens were repeated twice with same results. Protein expression in the APikL2 mutant screen was confirmed by western blot. Therefore, cell pellets of transformants were resuspended in SDS-PAGE loading buffer and incubated at 95°C for 10 min. After centrifugation at 13,000 rcf for 1 min supernatants were used for SDS-Page separation and western blotting. Membranes were probed with anti-GAL4-DNA-BD (Sigma) for the HMA domains in pGBKT7 and anti-GAL4-AD (Sigma) antibodies for the APikL effectors in pGADT7.

Expression and purification of proteins for in vitro binding studies and crystallization

To produce proteins for biophysical analyses, *Escherichia coli* (*E. coli*) SHuffle cells were transformed with goldengate compatible pOPIN expression vectors encoding sHMA94, sHMA25 and the AVR-PikL2 variants and mutants with a cleavable 6XHIS+protein GB1 domain (GB1) solubility tag under carbenicillin selection. Inoculated cell cultures were grown in autoinduction media [87] at 30°C for 6h and 18°C overnight and harvested by centrifugation at 5,500 g for 7 mins. Pelleted cells were resuspended in 50 mM HEPES pH 8.0, 500 mM NaCl, 30 mM imidazole, 50 mM glycine and 5% glycerol (buffer A1) supplemented with EDTA free protease inhibitor tablets. The cells were lysed by sonication with a VC 750 Vibra-Cell™ (Sonics) at 40% amplitude and cell debris was removed by centrifugation at 45,000 g for 45 mins. Initial purification of AVR-PikL2 or sHMA protein from the clarified lysate was performed by immobilised metal affinity chromatography (IMAC) through application of the clarified lysate to a Ni²⁺-NTA column connected to an AKTA Xpress purification system. 6xHis+GB1-AVR-PikL2 was step-eluted with elution buffer (buffer B1; buffer A containing 500 mM imidazole) and directly injected onto a Superdex 75 26/60 gel filtration column pre-equilibrated in 10 mM HEPES pH 8, 150 mM NaCl (buffer A4). Fractions containing 6xHis+GB1-AVR-PikL2 (as assessed by SDS-PAGE) were pooled and the 6xHis-GB1 tag was cleaved by addition of 3C protease (10 µg/mg fusion protein) with overnight incubation at 4°C. Cleaved AVR-PikL2 or sHMA protein was purified from the digest using a Ni²⁺-NTA column to perform reverse-IMAC, collecting the eluate and concentrating before a second round of size-exclusion. Proteins were then concentrated via 3 kDa cut-off spin concentrations to a final concentration of 10 mg/mL as calculated by absorbance at 205 by Nanodrop (ThermoFisher).

For crystallization of the sHMA/AVR-PikL2 complexes, AVR-PikL2A and AVR-PikL2F were cloned into pOPIN-GB1 vector mentioned above. sHMA proteins were cloned into goldengate compatible pOPIN vectors untagged under a kanamycin selection. *E. coli* SHuffle cells were then co-transformed with the pOPIN vectors encoding 6HIS-GB tagged AVR-PikL2 proteins and untagged sHMA proteins and then selected for on Kanamycin 30 µg / mL + Carbenicillin 100 µg / mL LB plates. Cell cultures were then inoculated and grown as mentioned above, under Kanamycin 100 µg / mL + Carbenicillin 100 µg / mL selection. Protein complexes were purified via IMAC coupled with size-exclusion as mentioned above. Untagged sHMA proteins were observed to be co-purified with the tagged effectors as assessed by SDS-PAGE. Confirmed protein complexes were concentrated to 10 mg/mL before storage at -80°C.

Crystallization, data collection and structure solution

For crystallization, AVR-PikL2/sHMA complexes were concentrated in buffer A4 (10 mM HEPES pH 7.4, 150 mM NaCl). Sitting drop, vapour diffusion crystallization trials were set up in 96-well plates, using an Oryx nano robot (Douglas Instruments). Crystallisation plates were incubated at 20°C.

AVR-PikL2F/sHMA94 crystals typically appeared after 4 days in SG1™ Screen (Molecular Dimensions), well D10 (0.2 M lithium sulfate, 0.1 M BIS-TRIS pH 6.5, 25% PEG3350). Crystals obtained from this condition were unsuitable for diffraction and were used for cross-seeding experiments. Cross-seeding into Morpheus HT-96 screen (Molecular Dimensions) resulted in formation of crystals suitable for diffraction in condition E3 (0.12 M ethylene glycols 0.1 M buffer system 1 pH 6.5, 30% v/v precipitant mix 3). AVR-PikL2A/sHMA25 crystals suitable for diffraction were harvested after 11 days from Morpheus (Molecular Dimensions) condition H10 (0.1 M amino acids, 0.1 M buffer system 3 pH 8.5, 50% precipitant mix 4). Crystals were snap frozen in liquid nitrogen prior to shipping to the Diamond Light Source.

X-ray data sets were collected at the Diamond Light Source. The AVR-PikL2F/sHMA94 dataset was collected on the i03 beamline under proposal mx18565 and the APikL2A/sHMA25 dataset was collected on the i04 beamline under proposal mx-13467. The data were processed using the xia2 pipeline and AIMLESS, as implemented in CCP4i2 [88]. The AVR-PikL2A/sHMA25 structure was solved by molecular replacement using PHASER [89] and the PikpHMA/AVR-PikD structure. The AVR-PikL2A/sHMA25 structure was then used to solve the AVR-PikL2F/sHMA94 structure via PHASER. The final structures were obtained through iterative cycles of manual rebuilding and refinement using COOT [90] and REFMAC5 [91], and ISODLE [92], as implemented in CCP4 and ChimeraX [93] respectively. Structures were validated using the tools provided in COOT and MOLPROBITY [94]. Protein interface analyses were performed using QtPISA [62]. For each complex, one AVR-PikL2 effector/sHMA protein assembly was used as a representative example. Models are visualised using PyMol 2.0 and ChimeraX. X-ray diffraction data can be found in the Protein Data Bank (<https://www.ebi.ac.uk/pdbe/>) with accession numbers 7NLJ (APikL2A/sHMA25) and 7NMM (APikL2F/sHMA94).

In vitro binding analysis with isothermal titration calorimetry

All calorimetry experiments were recorded using a MicroCal PEAQ-ITC (Malvern, UK). To test the interaction of AVR-PikL2 proteins with sHMA94, experiments were carried out in triplicate at 25°C using buffer A4. The calorimetric cell was filled with 60 μM AVR-PikL2 and titrated with 350 μM sHMA94 protein. For each ITC run, a single injection of 0.5 μL of ligand was followed by 19 injections of 2 μL each. Injections were made at 120-second intervals with a stirring speed of 750 rpm. Difficulty with measuring concentration of the protein due to extreme molar extinction coefficients ($\epsilon = 0.19$ and $2.6 \text{ M}^{-1}\text{cm}^{-1}$ at A_{280} for sHMA94 and AVR-PikL2 proteins, respectively) created difficulty in accurately titrating at an even stoichiometric ratio and therefore data was fit to a 1:1 binding model, with number of sites equal to one, as informed by the crystal structures of the complexes. Data were processed with AFFINImeter ITC analysis software [95]. Experimental replicates are shown in S6 Fig.

Homology modelling and conservation predictions

Homology models of APikL4, and APikL5 were generated using Phyre2 [96] One-to-One threading with AVR-PikD (PDB ID: 6G10 chain C) as a template. Conservation of residues in AVR-Pik family members was calculated using ConSurf [97] and visualized using PyMol.

Selection test and reconstruction of the evolutionary history of APiKL family members

Genomic sequences of all APiKL members were extracted from 107 *M. oryzae* genomes and multiple sequence alignments produced using MUSCLE (Edgar, 2004). Pseudogene sequences and positions with missing data (C-Terminal extension of APiKL5) were removed. We then generated a codon-based, maximum likelihood phylogenetic tree with 1000 bootstrap iterations of the family using MEGA 7 (Kumar et al., 2016). We then calculated the rates of dN/dS (ω) in the phylogenetic tree with two branch models implemented in codeml, the free-ratio model ($M = 1$; $NSsites = 0$) and the multi-ratio model ($M = 2$; $NSsites = 0$) allowing for separate ratios for each branch. The one-ratio model that allows only one dN/dS rate across all branches ($M = 0$; $NSsites = 0$) was rejected after performing likelihood ratio tests using the log likelihood values $\ln L_0 = -1849.510477$ and $\ln L_1 = -1829.104349$ (free ratio model) and $LR = 2 \times (\ln L_1 - \ln L_0) = 40.812256$ (1% significance level; degrees of freedom = 18). Additionally, we calculated the pairwise synonymous (dS) and nonsynonymous (dN) substitution rates using the codeml function YN00 of PAML v4.9 (54). To detect specific sites under positive selection, we used the site model implemented in PAML v4.9 by comparing the null models M0 (one-ratio), M1 (nearly neutral), M7 (beta), and three alternative models, M3 (selection), M2 (discrete), M8 (beta & ω) and performing likelihood ratios tests as described above. Model M3 and M8 passed the chi-square test at a significance level of 1% (S2 Table). Ancestral state reconstruction was performed by a codon based maximum likelihood analysis with codeml (included in the PAML v4.9 package) according to [53,54].

Supporting information

S1 Table. Isolates and genomic sequences used in this study.
(XLSX)

S2 Table. Functional annotation of B71 gene models at APiKL loci.
(XLSX)

S3 Table. Summary of the yeast two-hybrid interaction screen.
(XLSX)

S4 Table. Summary of X-ray data collection and refinement.
(DOCX)

S5 Table. Summary of the QtPISA interface analysis.
(XLSX)

S1 File. Fasta file of all APiKL amino acid sequences identified in this study.
(FAA)

S2 File Fasta file of all intact APiKL nucleotide sequences.
(FNA)

S1 Fig. The APiKL2 locus is conserved on chromosome 3 across host-specific lineages of *M. oryzae*. Whole genome alignments between the 70–15 reference genome assembly (Ensembl-Fungi assembly MG08) and seven isolates representative of five genetic lineages. Homologous chromosomes are color coded based on their best match in 70–15. Colored links between chromosomes show alignments >10 kb. Grey links show alignments in reverse direction (inversions). The location of each APiKL gene is indicated.
(TIF)

S2 Fig. Detailed assignment of electrostatic interactions in the APiK2A/sHMA25 and APiK2F/sHMA94 complexes. Cartoon representation of the two effector/sHMA complexes with ball and stick rendering of interacting residues as determined by PISA (EMBL-EBI). Hydrogen bonds are represented as a yellow dashed line with cut-off distance of 4 Å. A) Interface 1 B) Interface 2 C) Interface 3.
(TIF)

S3 Fig. The atomic interactions between Asp/Asn-66 and Gln-43/44 which facilitate the reorganisation of hydrogen bonding at the HMA binding interface. A, B) The hydrogen bonding networks of the APiK2A/sHMA25 and APiK2F/sHMA94 complexes around Gln-43/44, respectively. A) The OD1 atom of Asp-66 of APiK2 interacts with the NE2 atom of Gln-43 of sHMA25, preventing it from interacting with the OG1 atom and backbone carbonyl of Thr-37. B) The ND2 atom of Asn-66 interacts with the OE1 atom of Gln-44, allowing the NE2 atom of Gln-44 to interact with the OG1 atom and backbone carbonyl of Thr-38. C) Superimposition of the two APiK2/sHMA complexes. The blue and red arrows adjacent to Gln-43/44 indicate the movement of the NE2 and OE1 atoms between the two structures, exemplifying the rotation of the Gln-44 residue in the APiK2F/sHMA94 complex. Similarly, the red arrow adjacent to Thr-37/38 shows the movement of the OG1 atom in the APiK2F/sHMA94 complex, due to the interaction with the NE2 atom of Gln-44.
(TIF)

S4 Fig. Residues under diversifying selection locate to the loop regions and the N-terminal helix. Conservation scores (determined by ConSurf) of amino acid residues in the APiK-family mapped to the crystal structure of APiK2. Colors indicate the conservation score and confidence interval.
(TIF)

S5 Fig. The HMA-binding interface of APiK-family members is highly polymorphic. Schematic representation of the APiK family including all polymorphic sites. APiK effectors described in this study are highlighted in bold. The crucial polymorphism in APiK2 is highlighted in red. APiK family members in grey were not polymorphic in the population studied here. Asterisks indicate polymorphisms in the signal peptide.
(TIF)

S6 Fig. Experimental replicates of in vitro binding affinity measurements by ITC. The top panels show heat differences upon injection of the effector into the cell containing sHMA94. The second panel shows integrated heats of injection (•) and the best fit (red line) to a single site binding model calculated using AFFINImeter ITC analysis software. The final panel shows the difference between the fit to a single site binding model and the experimental data; the closer to zero indicates stronger agreement between the data and the fit.
(TIF)

Acknowledgments

We thank Clare Stevenson and Dave Lawson from the John Innes Centre protein X-ray crystallography platform for technical assistance during protein crystal preparation and data collection. We further want to thank all members of the BLASTOFF team at the Sainsbury Laboratory and John Innes Centre as well as Ki-Tae Kim and Yong-Hwan Lee for valuable discussions.

Author Contributions

Conceptualization: Mark J. Banfield, Sophien Kamoun, Thorsten Langner.

Data curation: Adam R. Bentham, Yohann Petit-Houdenot, Joe Win, Mark J. Banfield, Sophien Kamoun, Thorsten Langner.

Formal analysis: Adam R. Bentham, Yohann Petit-Houdenot, Joe Win, Izumi Chuma, Sophien Kamoun, Thorsten Langner.

Funding acquisition: Mark J. Banfield, Sophien Kamoun.

Investigation: Adam R. Bentham, Yohann Petit-Houdenot, Mark J. Banfield, Thorsten Langner.

Methodology: Joe Win, Mark J. Banfield, Thorsten Langner.

Project administration: Sophien Kamoun.

Resources: Izumi Chuma, Ryohei Terauchi, Thorsten Langner.

Supervision: Mark J. Banfield, Sophien Kamoun, Thorsten Langner.

Validation: Adam R. Bentham, Thorsten Langner.

Visualization: Adam R. Bentham, Sophien Kamoun, Thorsten Langner.

Writing – original draft: Adam R. Bentham, Sophien Kamoun, Thorsten Langner.

Writing – review & editing: Adam R. Bentham, Yohann Petit-Houdenot, Joe Win, Izumi Chuma, Ryohei Terauchi, Mark J. Banfield, Sophien Kamoun, Thorsten Langner.

References

1. Dong S, Stam R, Cano LM, Song J, Sklenar J, Yoshida K, et al. Effector Specialization in a Lineage of the Irish Potato Famine Pathogen. *Science*. 2014 Jan 31; 343(6170):552. <https://doi.org/10.1126/science.1246300> PMID: 24482481
2. Kim K-T, Ko J, Song H, Choi G, Kim H, Jeon J, et al. Evolution of the Genes Encoding Effector Candidates Within Multiple Pathotypes of *Magnaporthe oryzae*. *Front Microbiol*. 2019; 10:2575. <https://doi.org/10.3389/fmicb.2019.02575> PMID: 31781071
3. Kourelis J, Malik S, Mattinson O, Krauter S, Kahlon PS, Paulus JK, et al. Evolution of a guarded decoy protease and its receptor in solanaceous plants. *Nat Commun*. 2020 Sep 2; 11(1):4393. <https://doi.org/10.1038/s41467-020-18069-5> PMID: 32879321
4. Upton JL, Zess EK, Białas A, Wu C-H, Kamoun S. The coming of age of EvoMPMI: evolutionary molecular plant-microbe interactions across multiple timescales. *Curr Opin Plant Biol*. 2018 Aug; 44:108–16. <https://doi.org/10.1016/j.pbi.2018.03.003> PMID: 29604609
5. Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev*. 2015 Dec; 35:57–65. <https://doi.org/10.1016/j.gde.2015.09.001> PMID: 26451981
6. Cesari S. Multiple strategies for pathogen perception by plant immune receptors. *New Phytol*. 2018 Jul; 219(1):17–24. <https://doi.org/10.1111/nph.14877> PMID: 29131341
7. De la Concepcion JC, Franceschetti M, Maqbool A, Saitoh H, Terauchi R, Kamoun S, et al. Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen. *Nat Plants*. 2018 Aug 1; 4(8):576–85. <https://doi.org/10.1038/s41477-018-0194-x> PMID: 29988155
8. Kanzaki H, Yoshida K, Saitoh H, Fujisaki K, Hirabuchi A, Alaux L, et al. Arms race co-evolution of *Magnaporthe oryzae* AVR-Pik and rice Pik genes driven by their physical interactions: Co-evolution of fungal and rice genes. *Plant J*. 2012 Dec; 72(6):894–907. <https://doi.org/10.1111/j.1365-3113.2012.05110.x> PMID: 22805093
9. Kourelis J, Kamoun S. RefPlantNLR: a comprehensive collection of experimentally validated plant NLRs. *bioRxiv*. 2020 Jan 1;2020.07.08.193961.
10. Maqbool A, Saitoh H, Franceschetti M, Stevenson C, Uemura A, Kanzaki H, et al. Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife*. 2015 Aug 25; 4:e08709.

11. Chuma I, Isobe C, Hotta Y, Ibaragi K, Futamata N, Kusaba M, et al. Multiple Translocation of the AVR-Pita Effector Gene among Chromosomes of the Rice Blast Fungus *Magnaporthe oryzae* and Related Species. Tyler B, editor. PLoS Pathog. 2011 Jul 28; 7(7):e1002147. <https://doi.org/10.1371/journal.ppat.1002147> PMID: 21829350
12. Raffaele S, Win J, Cano LM, Kamoun S. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. BMC Genomics. 2010; 11(1):637. <https://doi.org/10.1186/1471-2164-11-637> PMID: 21080964
13. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Microbiol. 2012 Jun; 10(6):417–30. <https://doi.org/10.1038/nrmicro2790> PMID: 22565130
14. Yoshida K, Saunders DGO, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, et al. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. BMC Genomics. 2016 Dec; 17(1):370. <https://doi.org/10.1186/s12864-016-2690-6> PMID: 27194050
15. Chavarro-Carrero EA, Vermeulen JP, D E. Torres, Usami T, Schouten HJ, Bai Y, et al. Comparative genomics reveals the in planta-secreted *Verticillium dahliae* Av2 effector protein recognized in tomato plants that carry the V2 resistance locus. Environ Microbiol. 2020 Oct 19. <https://doi.org/10.1111/1462-2920.15288> PMID: 33078534
16. Houterman PM, Ma L, Van Ooijen G, De Vroomen MJ, Cornelissen BJC, Takken FLW, et al. The effector protein Avr2 of the xylem-colonizing fungus *Fusarium oxysporum* activates the tomato resistance protein I-2 intracellularly. Plant J. 2009 Jun 1; 58(6):970–8. <https://doi.org/10.1111/j.1365-3113X.2009.03838.x> PMID: 19228334
17. Li J, Cornelissen B, Rep M. Host-specificity factors in plant pathogenic fungi. Fungal Genet Biol. 2020 Nov 1; 144:103447. <https://doi.org/10.1016/j.fgb.2020.103447> PMID: 32827756
18. Sharma R, Mishra B, Runge F, Thines M. Gene Loss Rather Than Gene Gain Is Associated with a Host Jump from Monocots to Dicots in the Smut Fungus *Melanopsichium pennsylvanicum*. Genome Biol Evol. 2014 Jul 24; 6(8):2034–49. <https://doi.org/10.1093/gbe/evu148> PMID: 25062916
19. Sperschneider J, Gardiner DM, Thatcher LF, Lyons R, Singh KB, Manners JM, et al. Genome-Wide Analysis in Three *Fusarium* Pathogens Identifies Rapidly Evolving Chromosomes and Genes Associated with Pathogenicity. Genome Biol Evol. 2015 Jun 1; 7(6):1613–27. <https://doi.org/10.1093/gbe/evv092> PMID: 25994930
20. Stergiopoulos I, De Kock MJD, Lindhout P, De Wit PJGM. Allelic Variation in the Effector Genes of the Tomato Pathogen *Cladosporium fulvum* Reveals Different Modes of Adaptive Evolution. Mol Plant-Microbe Interactions. 2007 Sep 17; 20(10):1271–83. <https://doi.org/10.1094/MPMI-20-10-1271> PMID: 17918629
21. Thines M. An evolutionary framework for host shifts—jumping ships for survival. New Phytol. 2019 Oct 1; 224(2):605–17. <https://doi.org/10.1111/nph.16092> PMID: 31381166
22. Chiapello H, Mallet L, Guérin C, Aguilera G, Amselem J, Kroj T, et al. Deciphering Genome Content and Evolutionary Relationships of Isolates from the Fungus *Magnaporthe oryzae* Attacking Different Host Plants. Genome Biol Evol. 2015 Oct; 7(10):2896–912. <https://doi.org/10.1093/gbe/evv187> PMID: 26454013
23. Gladioux P, Condon B, Ravel S, Soanes D, Maciel JLN, Nhani A, et al. Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus *Magnaporthe oryzae*. mBio. 2018 Feb 27; 9(1):e01219–17. <https://doi.org/10.1128/mBio.01219-17> PMID: 29487238
24. Langner T, Białas A, Kamoun S. The Blast Fungus Decoded: Genomes in Flux. mBio. 2018 Apr 17; 9(2):e00571–18.
25. Thierry M, Milazzo J, Adreit H, Ravel S, Borrón S, Sella V, et al. Ecological Differentiation and Incipient Speciation in the Fungal Pathogen Causing Rice Blast. bioRxiv. 2020 Jan 1; 2020.06.02.129296.
26. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature. 2005 Apr; 434(7036):980–6. <https://doi.org/10.1038/nature03449> PMID: 15846337
27. Gómez Luciano LB, Tsai IJ, Chuma I, Tosa Y, Chen Y-H, Li J-Y, et al. Blast Fungal Genomes Show Frequent Chromosomal Changes, Gene Gains and Losses, and Effector Gene Turnover. Mol Biol Evol. 2019 Mar 5; 36(6):1148–61. <https://doi.org/10.1093/molbev/msz045> PMID: 30835262
28. Peng Z, Oliveira-Garcia E, Lin G, Hu Y, Dalby M, Migeon P, et al. Effector gene reshuffling involves dispensable mini-chromosomes in the wheat blast fungus. Lin X, editor. PLOS Genet. 2019 Sep 12; 15(9):e1008272. <https://doi.org/10.1371/journal.pgen.1008272> PMID: 31513573
29. Petit-Houdenot Y, Langner T, Harant A, Win J, Kamoun S. A Clone Resource of *Magnaporthe oryzae* Effectors That Share Sequence and Structural Similarities Across Host-Specific Lineages. Mol Plant-Microbe Interactions. 2020 May 27; 33(8):1032–5. <https://doi.org/10.1094/MPMI-03-20-0052-A> PMID: 32460610

30. Latorre SM, Reyes-Avila CS, Malmgren A, Win J, Kamoun S, Burbano HA. Differential loss of effector genes in three recently expanded pandemic clonal lineages of the rice blast fungus. *BMC Biol.* 2020 Jul 16; 18(1):88. <https://doi.org/10.1186/s12915-020-00818-z> PMID: 32677941
31. Tosa Y, Osue J, Eto Y, Oh H-S, Nakayashiki H, Mayama S, et al. Evolution of an Avirulence Gene, AVR1-CO39, Concomitant with the Evolution and Differentiation of *Magnaporthe oryzae*. *Mol Plant Microbe Interact.* 2005 Nov; 18(11):1148–60. <https://doi.org/10.1094/MPMI-18-1148> PMID: 16353550
32. Franceschetti M, Maqbool A, Jiménez-Dalmaroni MJ, Pennington HG, Kamoun S, Banfield MJ. Effectors of Filamentous Plant Pathogens: Commonalities amid Diversity. *Microbiol Mol Biol Rev.* 2017 Jun 1; 81(2):e00066–16. <https://doi.org/10.1128/MMBR.00066-16> PMID: 28356329
33. de Guillen K, Ortiz-Vallejo D, Gracy J, Fournier E, Kroj T, Padilla A. Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi. Xu J-R, editor. *PLOS Pathog.* 2015 Oct 27; 11(10):e1005228. <https://doi.org/10.1371/journal.ppat.1005228> PMID: 26506000
34. Cesari S, Thilliez G, Ribot C, Chalvon V, Michel C, Jauneau A, et al. The Rice Resistance Protein Pair RGA4/RGA5 Recognizes the *Magnaporthe oryzae* Effectors AVR-Pia and AVR1-CO39 by Direct Binding. *Plant Cell.* 2013 Apr 1; 25(4):1463. <https://doi.org/10.1105/tpc.112.107201> PMID: 23548743
35. Guo L, Cesari S, de Guillen K, Chalvon V, Mammri L, Ma M, et al. Specific recognition of two MAX effectors by integrated HMA domains in plant immune receptors involves distinct binding surfaces. *Proc Natl Acad Sci.* 2018 Nov 6; 115(45):11637. <https://doi.org/10.1073/pnas.1810705115> PMID: 30355769
36. Ortiz D, de Guillen K, Cesari S, Chalvon V, Gracy J, Padilla A, et al. Recognition of the *Magnaporthe oryzae* Effector AVR-Pia by the Decoy Domain of the Rice NLR Immune Receptor RGA5. *Plant Cell.* 2017 Jan 1; 29(1):156. <https://doi.org/10.1105/tpc.16.00435> PMID: 28087830
37. Zhang X, He D, Zhao Y, Cheng X, Zhao W, Taylor IA, et al. A positive-charged patch and stabilized hydrophobic core are essential for avirulence function of AvrPib in the rice blast fungus. *Plant J.* 2018 Oct 1; 96(1):133–46. <https://doi.org/10.1111/tpj.14023> PMID: 29989241
38. Zhang Z-M, Zhang X, Zhou Z-R, Hu H-Y, Liu M, Zhou B, et al. Solution structure of the *Magnaporthe oryzae* avirulence protein AvrPiz-t. *J Biomol NMR.* 2013 Feb 1; 55(2):219–23. <https://doi.org/10.1007/s10858-012-9695-5> PMID: 23334361
39. Bialas A, Zess EK, De la Concepcion JC, Franceschetti M, Pennington HG, Yoshida K, et al. Lessons in Effector and NLR Biology of Plant-Microbe Systems. *Mol Plant Microbe Interact.* 2018 Jan; 31(1):34–45. <https://doi.org/10.1094/MPMI-08-17-0196-FI> PMID: 29144205
40. Smith JR, Leong SA. Mapping of a *Magnaporthe grisea* locus affecting rice (*Oryza sativa*) cultivar specificity. *Theor Appl Genet.* 1994 Sep 1; 88(8):901–8. <https://doi.org/10.1007/BF00220794> PMID: 24186240
41. Valent B, Farrall L, Chumley FG. *Magnaporthe grisea* genes for pathogenicity and virulence identified through a series of backcrosses. *Genetics.* 1991 Jan; 127(1):87–101. <https://doi.org/10.1093/genetics/127.1.87> PMID: 2016048
42. Yoshida K, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Yoshida K, et al. Association Genetics Reveals Three Novel Avirulence Genes from the Rice Blast Fungal Pathogen *Magnaporthe oryzae*. *Plant Cell.* 2009 May 1; 21(5):1573. <https://doi.org/10.1105/tpc.109.066324> PMID: 19454732
43. Bialas A, Langner T, Harant A, Contreras MP, Stevenson CE, Lawson DM, et al. Two NLR immune receptors acquired high-affinity binding to a fungal effector through convergent evolution of their integrated domain. *bioRxiv.* 2021 Jan 1; 2021.01.26.428286. <https://doi.org/10.7554/eLife.66961> PMID: 34288868
44. De la Concepcion JC, Franceschetti M, MacLean D, Terauchi R, Kamoun S, Banfield MJ. Protein engineering expands the effector recognition profile of a rice NLR immune receptor. *eLife.* 2019 Sep 19; 8:e47713. <https://doi.org/10.7554/eLife.47713> PMID: 31535976
45. de Abreu-Neto JB, Turchetto-Zolet AC, de Oliveira LFB, Bodanese Zanettini MH, Margis-Pinheiro M. Heavy metal-associated isoprenylated plant protein (HIPP): characterization of a family of proteins exclusive to plants. *FEBS J.* 2013 Apr 1; 280(7):1604–16. <https://doi.org/10.1111/febs.12159> PMID: 23368984
46. Maidment JHR, Franceschetti M, Maqbool A, Saitoh H, Jantasuriyarat C, Kamoun S, et al. Multiple variants of the fungal effector AVR-Pik bind the HMA domain of the rice protein OsHIPP19, providing a foundation to engineer plant defence. *J Biol Chem.* <https://doi.org/https%3A//doi.org/10.1016/j.jbc.2021.100371> PMID: 33548226
47. Oikawa K, Fujisaki K, Shimizu M, Takeda T, Saitoh H, Hirabuchi A, et al. The blast pathogen effector AVR-Pik binds and stabilizes rice heavy metal-associated (HMA) proteins to co-opt their function in immunity. *bioRxiv.* 2020 Jan 1; 2020.12.01.406389.
48. Tehseen M, Cairns N, Sherson S, Cobbett CS. Metallochaperone-like genes in *Arabidopsis thaliana*. *Metallomics.* 2010; 2(8):556–64. <https://doi.org/10.1039/c003484c> PMID: 21072340

49. Fukuoka S, Saka N, Koga H, Ono K, Shimizu T, Ebana K, et al. Loss of Function of a Proline-Containing Protein Confers Durable Disease Resistance in Rice. *Science*. 2009 Aug 21; 325(5943):998–1001. <https://doi.org/10.1126/science.1175550> PMID: 19696351
50. Mirarab S, Reaz R, Bayzid MdS, Zimmermann T, Swenson MS, Warnow T. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*. 2014 Sep 1; 30(17):i541–8. <https://doi.org/10.1093/bioinformatics/btu462> PMID: 25161245
51. Mirarab S, Warnow T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*. 2015 Jun 15; 31(12):i44–52. <https://doi.org/10.1093/bioinformatics/btv234> PMID: 26072508
52. Sayyari E, Mirarab S. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Mol Biol Evol*. 2016 Jul 1; 33(7):1654–68. <https://doi.org/10.1093/molbev/msw079> PMID: 27189547
53. Langner T, Harant A, Gomez-Luciano LB, Shrestha RK, Malmgren A, Latorre SM, et al. Genomic rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast fungus. *PLOS Genet*. 2021 Feb 16; 17(2):e1009386. <https://doi.org/10.1371/journal.pgen.1009386> PMID: 33591993
54. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*. 1997 Oct 1; 13(5):555–6. <https://doi.org/10.1093/bioinformatics/13.5.555> PMID: 9367129
55. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007 Aug; 24(8):1586–91. <https://doi.org/10.1093/molbev/msm088> PMID: 17483113
56. Bos JIB, Chaparro-Garcia A, Quesada-Ocampo LM, Gardener BBM, Kamoun S. Distinct Amino Acids of the *Phytophthora infestans* Effector AVR3a Condition Activation of R3a Hypersensitivity and Suppression of Cell Death. *Mol Plant-Microbe Interactions*. 2009 Feb 9; 22(3):269–81. <https://doi.org/10.1094/MPMI-22-3-0269> PMID: 19245321
57. Daverdin G, Rouxel T, Gout L, Aubertot J-N, Fudal I, Meyer M, et al. Genome Structure and Reproductive Behaviour Influence the Evolutionary Potential of a Fungal Phytopathogen. *PLOS Pathog*. 2012 Nov 8; 8(11):e1003020. <https://doi.org/10.1371/journal.ppat.1003020> PMID: 23144620
58. Liu Z, Bos JIB, Armstrong M, Whisson SC, da Cunha L, Torto-Alalibo T, et al. Patterns of Diversifying Selection in the Phytotoxin-like scr74 Gene Family of *Phytophthora infestans*. *Mol Biol Evol*. 2005 Mar 1; 22(3):659–72. <https://doi.org/10.1093/molbev/msi049> PMID: 15548752
59. Oh S-K, Young C, Lee M, Oliva R, Bozkurt TO, Cano LM, et al. In Planta Expression Screens of *Phytophthora infestans* RXLR Effectors Reveal Diverse Phenotypes, Including Activation of the *Solanum bulbocastanum* Disease Resistance Protein Rpi-blb2. *Plant Cell*. 2009 Sep 1; 21(9):2928. <https://doi.org/10.1105/tpc.109.068247> PMID: 19794118
60. Oliva RF, Cano LM, Raffaele S, Win J, Bozkurt TO, Belhaj K, et al. A Recent Expansion of the RXLR Effector Gene Avrblb2 Is Maintained in Global Populations of *Phytophthora infestans* Indicating Different Contributions to Virulence. *Mol Plant-Microbe Interactions*. 2015 Apr 20; 28(8):901–12. <https://doi.org/10.1094/MPMI-12-14-0393-R> PMID: 25894205
61. Van de Wouw AP, Cozijnsen AJ, Hane JK, Brunner PC, McDonald BA, Oliver RP, et al. Evolution of Linked Avirulence Effectors in *Leptosphaeria maculans* Is Affected by Genomic Environment and Exposure to Resistance Genes in Host Plants. *PLOS Pathog*. 2010 Nov 4; 6(11):e1001180. <https://doi.org/10.1371/journal.ppat.1001180> PMID: 21079787
62. Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. *J Mol Biol*. 2007 Sep 21; 372(3):774–97. <https://doi.org/10.1016/j.jmb.2007.05.022> PMID: 17681537
63. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016 Jul 8; 44(W1):W344–50. <https://doi.org/10.1093/nar/gkw408> PMID: 27166375
64. Breen S, Williams SJ, Winterberg B, Kobe B, Solomon PS. Wheat PR-1 proteins are targeted by necrotrophic pathogen effector proteins. *Plant J*. 2016 Oct 1; 88(1):13–25. <https://doi.org/10.1111/tbj.13228> PMID: 27258471
65. Hogenhout SA, Van der Hoorn RAL, Terauchi R, Kamoun S. Emerging Concepts in Effector Biology of Plant-Associated Organisms. *Mol Plant-Microbe Interactions*. 2009 Jan 8; 22(2):115–22. <https://doi.org/10.1094/MPMI-22-2-0115> PMID: 19132864
66. Win J, Chaparro-Garcia A, Belhaj K, Saunders DGO, Yoshida K, Dong S, et al. Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb Symp Quant Biol*. 2012; 77:235–47. <https://doi.org/10.1101/sqb.2012.77.015933> PMID: 23223409
67. Maidment JHR, Franceschetti M, Maqbool A, Saitoh H, Jantasuriyarat C, Kamoun S, et al. Multiple variants of the blast fungus effector AVR-Pik bind the HMA domain of the rice protein OsHIPP19 with high affinity. *bioRxiv*. 2020 Jan 1; 2020.12.01.403451.

68. Inoue Y, Vy TTP, Yoshida K, Asano H, Mitsuoka C, Asuke S, et al. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. *Science*. 2017; 357(6346):80–3. <https://doi.org/10.1126/science.aam9654> PMID: 28684523
69. Chung H, Goh J, Han S-S, Roh J-H, Kim Y, Heu S, et al. Comparative Pathogenicity and Host Ranges of *Magnaporthe oryzae* and Related Species. *Plant Pathol J*. 2020 Aug 1; 36(4):305–13. <https://doi.org/10.5423/PPJ.FT.04.2020.0068> PMID: 32788889
70. Islam MT, Croll D, Gladieux P, Soanes DM, Persoons A, Bhattacharjee P, et al. Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *BMC Biol*. 2016 Dec; 14(1):84. <https://doi.org/10.1186/s12915-016-0309-7> PMID: 27716181
71. Islam MT, Kim K-H, Choi J. Wheat Blast in Bangladesh: The Current Situation and Future Impacts. *Plant Pathol J*. 2019 Feb 1; 35(1):1–10. <https://doi.org/10.5423/PPJ.RW.08.2018.0168> PMID: 30828274
72. Malaker PK, Barma NCD, Tiwari TP, Collis WJ, Duveiller E, Singh PK, et al. First Report of Wheat Blast Caused by *Magnaporthe oryzae* Pathotype triticum in Bangladesh. *Plant Dis*. 2016 Jun 6; 100(11):2330–2330.
73. Tembo B, Mulenga RM, Sichilima S, M'siska KK, Mwale M, Chikoti PC, et al. Detection and characterization of fungus (*Magnaporthe oryzae* pathotype Triticum) causing wheat blast disease on rain-fed grown wheat (*Triticum aestivum* L.) in Zambia. *PLOS ONE*. 2020 Sep 21; 15(9):e0238724. <https://doi.org/10.1371/journal.pone.0238724> PMID: 32956369
74. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci*. 2010 Oct 1; 35(10):539–46. <https://doi.org/10.1016/j.tibs.2010.04.009> PMID: 20541943
75. Gohlke H, Kuhn LA, Case DA. Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins Struct Funct Bioinforma*. 2004 Aug 1; 56(2):322–37. <https://doi.org/10.1002/prot.20116> PMID: 15211515
76. Dunbrack RL, Karplus M. Backbone-dependent Rotamer Library for Proteins Application to Side-chain Prediction. *J Mol Biol*. 1993 Mar 20; 230(2):543–74. <https://doi.org/10.1006/jmbi.1993.1170> PMID: 8464064
77. Liang S, Grishin NV. Side-chain modeling with an optimized scoring function. *Protein Sci*. 2002 Feb 1; 11(2):322–31. <https://doi.org/10.1110/ps.24902> PMID: 11790842
78. Fukuhara N, Kawabata T. HOMCOS: a server to predict interacting protein pairs and interacting sites by homology modeling of complex structures. *Nucleic Acids Res*. 2008 Jul 1; 36(Web Server issue):W185–189. <https://doi.org/10.1093/nar/gkn218> PMID: 18442990
79. Launay G, Simonson T. Homology modelling of protein-protein complexes: a simple method and its possibilities and limitations. *BMC Bioinformatics*. 2008 Oct 9; 9(1):427. <https://doi.org/10.1186/1471-2105-9-427> PMID: 18844985
80. Seong K, Krasileva KV. Computational structural genomics unravels common folds and predicted functions in the secretome of fungal phytopathogen *Magnaporthe oryzae*. *bioRxiv*. 2021 Jan 1; 2021.01.25.427855.
81. Tanaka S, Schweizer G, Rössel N, Fukada F, Thines M, Kahmann R. Neofunctionalization of the secreted Tin2 effector in the fungal pathogen *Ustilago maydis*. *Nat Microbiol*. 2019 Feb 1; 4(2):251–7. <https://doi.org/10.1038/s41564-018-0304-6> PMID: 30510169
82. Zess EK, Jensen C, Cruz-Mireles N, De la Concepcion JC, Sklenar J, Stephani M, et al. N-terminal β -strand underpins biochemical specialization of an ATG8 isoform. *PLOS Biol*. 2019 Jul 22; 17(7):e3000373. <https://doi.org/10.1371/journal.pbio.3000373> PMID: 31329577
83. Delaux P-M, Schornack S. Plant evolution driven by interactions with symbiotic and pathogenic microbes. *Science*. 2021 Feb 19; 371(6531):eaba6605. <https://doi.org/10.1126/science.aba6605> PMID: 33602828
84. Langner T, Kellner R, Harant A, Terauchi R, Kamoun S A clone resource of cereal heavy-metal associated (HMA) domains—putative targets of plant pathogen effector proteins. 2021 Jul 30 <https://zenodo.org/record/5148559>
85. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004 Mar 1; 32(5):1792–7. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147
86. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016 Jul; 33(7):1870–4. <https://doi.org/10.1093/molbev/msw054> PMID: 27004904
87. Studier FW. Stable Expression Clones and Auto-Induction for Protein Production in *E. coli*. *Structural Genomics: General Applications*. Totowa, NJ: Humana Press; 2014. p. 17–32. https://doi.org/10.1007/978-1-62703-691-7_2

88. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the *ccp4* suite and current developments. *Acta Crystallogr Sect D*. 2011 Apr; 67(4):235–42. <https://doi.org/10.1107/S0907444910045749> PMID: 21460441
89. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. *Phaser* crystallographic software. *J Appl Crystallogr*. 2007 Aug; 40(4):658–74. <https://doi.org/10.1107/S0021889807021206> PMID: 19461840
90. Emsley P, Cowtan K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr Sect D*. 2004 Dec; 60(12 Part 1):2126–32. <https://doi.org/10.1107/S0907444904019158> PMID: 15572765
91. Murshudov GN, Vagin AA, Dodson EJ. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr Sect D*. 1997 May; 53(3):240–55. <https://doi.org/10.1107/S0907444996012255> PMID: 15299926
92. Croll TI. *ISOLDE*: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr Sect D*. 2018 Jun; 74(6):519–30. <https://doi.org/10.1107/S2059798318002425> PMID: 29872003
93. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*. 2021 Jan 1; 30(1):70–82. <https://doi.org/10.1002/pro.3943> PMID: 32881101
94. Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D*. 2010 Jan; 66(1):12–21. <https://doi.org/10.1107/S0907444909042073> PMID: 20057044
95. Piñeiro Á, Muñoz E, Sabín J, Costas M, Bastos M, Velázquez-Campoy A, et al. *AFFINImeter*: A software to analyze molecular recognition processes from experimental data. *Anal Biochem*. 2019 Jul 15; 577:117–34. <https://doi.org/10.1016/j.ab.2019.02.031> PMID: 30849378
96. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The *Phyre2* web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015 Jun 1; 10(6):845–58.
97. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, et al. *ConSurf 2005*: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res*. 2005 Jul 1; 33(suppl_2):W299–302. <https://doi.org/10.1093/nar/gki370> PMID: 15980475