Structure-guided secretome analysis of gall-forming microbes offers insights into effector diversity and evolution

Soham Mukhopadhyay[1-3]*, Muhammad Asim Javed[1-3], Jiaxu Wu[1-3] and Edel Pérez-López[1-3]*

[1]Départment de Phytologie, Faculté des sciences de l'agriculture et de l'alimentation, Université Laval, Quebec City, Quebec, Canada.

[2]Centre de recherche et d'innovation sur les végétaux (CRIV), Université Laval, Quebec City, Quebec, Canada.

[3]L'Institute EDS, Université Laval, Quebec City, Québec City, Québec, Canada

*Corresponding: somuk3@ulaval.ca, edel.perez-lopez@fsaa.ulaval.ca

ABSTRACT

Phytopathogens secrete effector molecules to manipulate host immunity and metabolism. Recent advances in structural genomics have identified fungal effector families whose members adopt similar folds despite sequence divergence, highlighting their importance in virulence and immune evasion. To extend the scope of comparative structure-guided analysis to more evolutionarily distant phytopathogens with similar lifestyles, we used AlphaFold2 to predict the 3D structures of the secretome from selected plasmodiophorid, oomycete, and fungal gall-forming pathogens. Clustering protein folds based on structural homology revealed species-specific expansions and a low abundance of known orphan effector families. We identified novel sequence-unrelated but structurally similar (SUSS) effector clusters, rich in conserved motifs such as 'CCG' and 'RAYH'. We demonstrate that these motifs likely play a central role in maintaining the overall fold. We also identified a SUSS cluster adopting a nucleoside hydrolase-like fold conserved among various gall-forming microbes. Notably, ankyrin proteins were significantly expanded in gall-forming plasmodiophorids, with most being highly expressed during clubroot disease, suggesting a role in pathogenicity. Altogether, this study advances our understanding of secretome landscapes in gall-forming microbes and provides a valuable resource for broadening structural phylogenomic studies across diverse phytopathogens.

KEYWORDS: Effector biology, AlphaFold, Plasmodiophorid, Gall-forming pathogens, SUSS

INTRODUCTION

The evolutionary arms race between host plants and their microbial pathogens is a fascinating example of adaptation and counteradaptation. Central to this battlefield are the effectors — secreted proteins from pathogens that manipulate host cellular processes to facilitate infection and colonization [1]. Effectors can also be recognized by plant receptor and trigger immunity and defense responses in the host [2]. Due to their central role in pathogen-host interactions, effectors must continually evolve to evade detection by the host's immune system [3]. This arms race drives rapid changes in effector sequences, often resulting in high mutation rates and diversification [4]. Additionally, effectors should maintain structural integrity for functionality while altering surface residues to avoid immune recognition [5]. Recent studies have defined these groups of effectors as Sequence-Unrelated Structurally Similar (SUSS), which despite lacking sequence similarity, share significant structural homology [6,7]. For example, the MAX effector family, which includes proteins from various fungal pathogens like *Magnaporthe oryzae* and *Pyrenophora tritici-repentis*, exhibits a conserved β-sandwich fold[8]. Other examples include the RXLR-WY effector families in oomycetes [9,10], the LARS effectors in *Cladosporium fulvum* and *Leptosphaeria maculans* [11], the RALPH effectors in *Blumeria graminis* [12], and the FOLD effectors in *Fusarium oxysporum f. sp. lycopersici* (D. S. Yu et al., 2024) demonstrating structural conservation. FOLD effectors have also recently been found in the secretome of unrelated pathogenic and symbiotic fungi, pointing to the fold's relevance in plant colonization and expansion in different evolutive groups [14]. A recent study that classified orphan effector candidates (OECs) from Ascomycota into 62 main structural groups, proposes that such structural conservation can be explained by changes in thermodynamic frustration at surface residues, which increase the robustness of the protein structure while altering potential interaction sites [5].

Some of the discussed studies have been fueled by the emergence of machine-learning tools like AlphaFold [15], which has revolutionized the field of protein modeling and enabled the computational prediction of pathogen effector structures. The utility of AlphaFold in plant-microbe interactions has been further demonstrated by its ability to predict the structure of the highly conserved AvrE-family of effector proteins [16]. These proteins are crucial in the pathogenesis of various phytopathogenic bacteria but are challenging to study due to their large size, toxicity to plant and animal cells, and low sequence similarity to known proteins [17]. The structure prediction revealed β-barrel structures similar to bacterial porins, allowing to modulate host cell functions by facilitating the movement of small molecules across plasma membrane. AlphaFold Multimer [18], an

extension of AlphaFold designed to predict in-silico protein-protein interactions, has been recently used to identify 15 effector candidates capable of targeting the active sites of chitinases and proteases [19], expanding the applicability of these tools to advance the tailored functional characterization of effector proteins.

Recent advances in structure-guided secretome analysis have mostly focused on fungal pathogens due to their significant economic impact and the availability of robust genomic and transcriptomic resources. However, in recent years, there is a growing interest in protist pathogens, which are quickly becoming a threat to agriculture and environment [20]. For example, the gall-forming *Plasmodiophora brassicae*, a protist belonging to the class Plasmodiophorid, can cause significant yield loss in canola fields [21,22]. Moreover, these obligate biotrophic protists are impossible to culture in axenic media and therefore difficult to transform [23]. Although the effector repertory for some of these protists has been predicted, the majority of their secretome remains uncharacterized due to the absence of known protein domains [20]. To gain more insights into the secretome composition and effector biology of these understudied pathogens, we conducted structural similarity-based clustering of the predicted effectors of selected plasmodiophorid, oomycete, and fungal gall-farming pathogens. Here we examined (*i*) if the primary secretome families present in each pathogen share common folds with known fungal effector families; (*ii*) if they share common folds that could be associated to their biotrophic lifestyle and gall- forming pathogenicity strategy; and (*iii*) if some of the known effectors from these pathogens are part of SUSS effector families. By comparing the secretome of gall-forming pathogens from distant lineages, we provide a comprehensive overview of the uniqueness and commonality of the secretome landscape and offers more insights into the protist effector families by bringing them into the structural genomics era.

RESULTS

Secretome prediction and structural modeling of gall forming pathogens

Based on genome availability, phylogenetic distance, and economic significance, six gall-forming pathogens were selected for secretome analysis- two plasmodiophorids, *Plasmodiophora brassicae* and *Spongospora subterranea*, oomycete *Albugo candida*, and three fungi from different lineages, *Taphrina deformans, Ustilago maydis*, and *Synchytrium endobioticum* (**Fig. 1a**). To gain better understanding of plasmodiophorids, for which structural data is scarce, we also included *Polymyxa betae*, a non-gall-forming plasmodiophorid vector of the beet necrotic yellow vein virus,

causing Rhizomania disease [24] (**Fig. 1a**). To identify the putative secretome of these plant pathogens, we first employed SignalP [25] to predict sequences with N-terminal signal peptide. Sequences carrying a signal peptide were subjected to DeepTMHMM [26] search to remove sequences with transmembrane domains. Mature protein sequences greater than 1000 amino acid length were also removed (**Fig. 1b**). The *Ustilago maydis* secretome and corresponding structures were obtained from a recent study [7] using similar filtering steps. This resulted in a total of 4197 proteins from seven gall-forming or related plant pathogens (**Table S1**). Next, the structures of 3575 proteins, excluding the *U. maydis* secretome which was already available, were modeled using AlphaFold2. InterproScan [27] was performed on all 4197 proteins against pfam, Gene3D and SUPERFMAILY databases, identifying 41.59% of the proteins analyzed carrying a known protein domain (**Table S1**). Next, 2349 structures with pLDDT > 65 were selected for further analysis (**Table S1**). Although pLDDT > 70 is recommended as the threshold for structures with reliable confidence by AlphaFold developers, we selected a score of 65 or higher to include *AvrSen1* (pLDDT 67), the only known *S. endobioticum* avirulent gene [28] (**Fig. 1c**). This resulted in the inclusion of 298 candidate effectors which would have otherwise been excluded (**Table S1**). IUPred3 [29] was also used to score the proteins as disordered or not based on whether 50% of the sequence position was predicted to be disordered (Table **S1**). Most of the proteins (97%) with pLDDT > 70 were not disordered, while 26% of the secretome with pLDDT < 70 was disordered, thus showing the limitation of Alpahfold2 in modelling such effectors (**Table S1**).

Structure-based clustering reveals species-specific folds and low homology with known fungal effector families

Structural similarity among the predicted structures were assessed using TMAlign [30] by scoring all structures against each other. To study the similarity with known effector families, we also included 19 crystal structures from the DELD, FOLD, LARS, MAX, KP6, RALPH, NTF2-like, ToxA, C2-like, and Zn-binding effector families, as previously utilized in a recent study [14] (**Table S2**). Additionally, we also included 62 structural families of orphan candidate effectors (OCE) recently identified in the Ascomycota lineage [5] (**Table S2**). Comparisons with TMScore greater than 0.5 were considered positive for structural similarity. Markov clustering [31,32] with an inflation value of two was applied to cluster the secretome based on structural similarity score. This resulted in 254 structural clusters with at least two members (**Fig. 2**, **Table S3**).

Ankyrin repeat containing proteins were the largest cluster for plasmodiophorid *P. brassicae* (n=42) and *S. subterranea* (n=39) (**Fig. 2**). The largest cluster detected on *A. candida* secretome is formed by 'CCG' [33] class of effectors (n=48), while *U. maydis* largest cluster was composed of the Tin2-like proteins [34] (n=31). For *S. synchytrium*, the largest cluster (n=64) contains AvrSen1 virulence factor (**Fig. 2a**). *P. betae* and *T. deformans* do not carry large (n > 30) effector clusters which could be due to *P. betae's* vector-like nature and *T. deformans*'s reduced genome size compared to other fungi [35]. *T. deformans*'s largest cluster is primarily composed of various Glycoside hydrolases while *P. betae*'s largest cluster consists of orphan helical proteins (**Table S3**). A 20-member Kinase family (cluster 13) was noted in *P. brassicae* which was absent in *A. candida* and 3 other fungi. Chitin deacetylases, which have been reported to convert chitin to less immunogenic chitosan [36] was expanded (n=11) in *P. brassicae* (**Table S3**). Chitin deacetylases were found in 5 out of 7 pathogens tested, except in oomycete *A. candida* and ascomycete *T. deformans*, which despite being a fungus contains very little chitin in the cell wall [37] (**Table S3**). In *S. subterranea* and *P. brassicae*, a unique TauD/TfdA protein cluster was also identified, typically found in bacteria for taurine utilization as a sulfur source [38]. *P. betae* carries a 6 member G-domain protein cluster, with similarity to it's host *Beta vulgari*'s GTPases (**Fig. S1**, **Table S3**). Among the well characterized fungal effector families, only three KP6 fold [39] and one RALPH like fold were found in *U. maydis* (**Fig. S1**, **Table S3**). *T. deformans* carry 2 ToxA homologs (**Fig. S1**, **Table S3**). .

Effector folds conserved across kingdoms

Six protein folds—Hydrolases (clusters 2, 8, 9), Carboxypeptidases (cluster 12), Aspartyl proteases (cluster 3), Lectins (cluster 38), SCP domain (cluster 24), and an orphan group (cluster 5)—contained proteins from all the pathogens investigated in this study, indicating deep evolutionary conservation of these folds across kingdoms (**Table S3**). In fact, 20 out of 62 Ascomycete orphan effector groups had at least one structural homolog in plasmodiophorids and oomycetes tested in this study, although a complete evolutionary connection would require the comparison of a much larger number of pathogens (**Table S3**). We didn't find any specific fold (n > 1 per species) conserved only among the gall-forming pathogens studied, indicating that this virulence strategy can be achieved by different mechanisms without necessarily converging onto common effector folds.

Structural search identifies a nucleoside hydrolase-like fold conserved in some gall-forming pathogens

Effectors are notorious for carrying unknown domains, making it difficult to predict the putative function of promising candidates [41]. We searched for uncharacterized *P. brassicae* candidate effectors within the same cluster that were also overexpressed during infection. Two candidate effectors, PBTT_09143 and PBTT_07479, which are the first and fifth most expressed proteins at 16 days post inoculation (dpi) during clubroot disease, belong to the cluster 21 grouping with PBTT_0412, none of which carries a known domain (**Fig. 3a-c**). Interestingly, the cluster also includes ten members from *A. candida* secretome, with eight carrying a predicted nucleoside hydrolase domain (**Table S4**). We subjected the three *P. brassicae* candidates to a FoldSeek-mediated [42] structural search against the PDB100 and AFD-proteome databases. Nucleoside hydrolases always emerged as the top hit (E-value < 10^-5, Prob ~1, TMscore >=0.5) (**Table S4**). Next, searching AFDB cluster webtool [43], which allows for the identification of structural homologs across the known protein space, AFDB clusters A0A0G4IP88 and A0A024FV66 emerged as the top hits. The members of these clusters are often gall-forming, carry predicted signal peptide and belong to various biotrophs *like Melanopsichium pennsylvanicum, Ustilago maydis, Albugo candida, Sporisorium scitamineum, Testicularia cyperi, Colletotrichum orbiculare,* among others (**Fig 3b, d**). Some of the members also do not carry identifiable domains (**Table S4**) and show limited sequence similarity among themselves (**Fig 3b**, **Table S4**). Interestingly, the cluster also includes bacterial Type III effector HopQ1 from *Pseudomonas syringae* (**Table S4**). Unlike *P. brassicae* effectors, HopQ1 is predicted to carry a nucleoside hydrolase domain, although the domain has been reported to be unable to bind standard nucleosides [44]. HopQ1 has been reported to be associated with 14-3-3 plant proteins to promote virulence [44]. Thus, it's possible that the nucleoside hydrolase-like fold present in various gall-forming fungal, protist, and oomycete pathogens might have also neo-functionalized and be involved in new molecular strategies during the infection.

A fungal effector family shows structural homology despite extreme sequence diverge

The Mig1 protein in *Ustilago maydis* is a maize-induced effector that plays a crucial role in the biotrophic interaction between the fungus and its host [45]. It is specifically induced during the biotrophic phase, contributing to the fungus's pathogenicity. Cluster 30, specifically found in *Ustilago maydis*, contains this effector (**Table S5**). Upon examining the sequence and structural

similarities between the members of the cluster, we discovered instances of structural homology (TMScore > 0.5) despite pairwise sequence identity being as low as 0.6% and no higher than 30% (**Fig. 4a-d**, **Table S5**). When aligning the protein sequences of 13 members, four cysteine residues were found to be strongly conserved (**Fig. 4e**). These four residues form two disulfide bridges (**Fig. 4f**), likely playing a crucial role in maintaining the overall fold despite significant sequence dissimilarity. All 13 members of the cluster were expressed upon infection (**Table S6**). A variable expression was observed for the *Mig1*-like genes located in the same genomic region, including members with completely opposite patterns of induction (**Table S6**), suggesting a possible regulatory role or the acquisition of new functions.

Identification of new SUSS effector families enriched in known motifs

The fungi *S. endobioticum* and *A. candida* encode large effector clusters which included previously identified avirulent factors AvrSen1 and CCG28/31/33/70, respectively (**Table S7**) [46]. It was previously shown that the N-terminal region of the CCG28 ,33 and 70 share structural homology [46]. To examine if these clusters represent SUSS families, we carried out the sequence-based clustering of the 4197 proteins from the seven pathogens investigated here. We performed BlastP search in all-vs-all mode and kept only those results with E-value lower than 10^-04 and bidirectional coverage of 50%. Markov clustering revealed 642 sequence-based clusters with at least two members (**Table S7**). We searched for sequence clusters having members from the same structural clusters. This revealed the presence of 12 sequence-related clusters associated with AvrSen1 structural cluster and 11 sequence-related clusters associated with CCG structural cluster (**Fig. 5a**, **Table S7**). It was previously reported that CCG containing effectors share limited similarity around the CCG motif and can be grouped in several clades based on sequence similarity [33]. Thus, AvrSen1 and CCG represent novel SUSS effector families whose similarities can’t be delineated by sequence search alone. Integration of sequence and structure data increased the member count of AvrSen1 and CCG clusters to 124 and 50 from previous 64 and 48, respectively (**Table S8**).

It has been reported that *S. endobioticum* secretome is enriched in RAYH [47].While it is understood that the CCG class of effectors derived its name due to the presence of the CCG motif, it was not immediately clear if Avrsen1 cluster was also the source of the conserved RAYH motif. To verify that, we subjected the mature protein sequences of the two pathogen secretome to motif

search using MEME [48]. Here we identified a 16 amino acid long RAYH motif present in 118 *S. endobioticum* proteins and 15 amino acid long 'CCG' motif present in 74 *A. candida* proteins (E < 0.1, combined match p < 0.001) (**Fig. 5b**, **Fig. S2, S3**, **Table S8**). Of the 118 sequences proteins carrying the RAYH motif, 79 were members of the AvrSen1 cluster in *S. endobioticum* (**Table S8**). Thus, the expansion of SUSS effectors in *A. candida* and *S. synchytrium* has resulted in the enrichment of common motifs, something that has recently been observed for the Y/F/WxC motif in *Blumeria graminis* RNA-like effector cluster [7].

Selection pressure analysis on *A. candida* CCG members shows that both cysteine residues in the 'CCG' motifs and two additional cysteine residues within 50 amino acids of the motif are often under purifying selection. (**Fig. S4, S5**). Visualizing these four cysteines on the predicted structure shows that they form disulfide bridges and probably play a crucial role in overall maintenance of the fold (**Fig. 5c**). The CCG motif seems to be a crucial part of a module consisting of two parallel alpha-helices joined to a beta-sheet and CCG effectors are often composed of several of these modules (**Fig. 5c**). The RAYH motif was also found to be part of the core structure of most AvrSen1-like effectors, forming a long alpha-helix (**Fig 5d**). Apart from the RAYH motif region, several surrounding hydrophobic residues were also strongly conserved, likely playing a role in maintaining the structure (**Fig. S6**). Examining the sequence-related subclusters of the AvrSen1-like family we found that the effectors are evolving by keeping the core alpha-helix bundle fixed while diversifying the peripheral stretches (**Fig 5d**).

Ankyrin repeat containing proteins are a common feature of gall forming plasmodiophorids

The largest structural cluster in *P. brassicae* and *S. subterranea* consists of ankyrin repeat-containing proteins (hereafter ANK proteins) (**Fig. 6a**). The presence of only five members in the non-gall-forming plasmodiophorid *P. betae*, in contrast to over 40 members in the gall-forming *P. brassicae* and *S. subterranea*, highlights the importance of this domain in their pathogenicity strategies (**Table S9**). Interestingly, InterProScan identified 32 additional *P. brassicae* proteins with ANK domains, while only two *S. subterranea* ANK proteins could be identified outside of the cluster (**Table S9**). We found that *P. brassicae* secretome is richer in repeat proteins compared to *S. subterranean* ANK proteins repertory (**Fig 6a**), with 19 additional leucine-rich repeats (LRRs), ten tetratricopeptide repeats, and three MORN repeats (**Table S9**). Although 40 out of 74 ankyrins are over expressed (TPM >10) at 16- and 26 dpi in *Arabidopsis thaliana*, only five LRRs are induced

(**Table S9**). Notably, 17 ankyrin-repeat proteins also carry the SKP1/BTB/POZ domain, which is often involved in ubiquitination [49] (**Fig. 6b**, **Table S9**).

A MEME motif scan with the members of the ankyrin cluster identified a 33 amino acid long motif in *P. brassicae* and a 32 amino acid length motif in *S. spongospora* (**Fig 6c**, **Table S9**). Aligning the MEME profile of the two identified ankyrin motifs shows a strong conservation of two Leucine and one Alanine residues (**Fig. 6d**) and upon visualization, those residues form the hydrophobic pocket between the two alpha helices of the ankyrin repeat, stabilizing the structure (**Fig 6d**). The rest of the non-conserved residues were found to be highly polymorphic (**Fig. 6d**).

Finally, to have an idea of the possible role of ANK proteins in plasmodiophorid virulence, we selected 70 ankyrin domain-containing proteins from *P. brassicae* and screened them against 20 key immune-related genes in *Arabidopsis* using AlphaFold-Multimer (**Table S10**). Protein-protein interactions were considered significant if the inter-chain Predicted Aligned Error (PAE) value was below 10, and the iPTM+pTM score was 70 or higher. Among the identified interactions, MPK3, MAPK4, MAPK6 SnRK1, NPR1, XCP1, CNGC4, and BAK1 were targeted by a total of ten ankyrin domain proteins (**Fig. 6e**, **Table S11**). This dataset should serve as a valuable starting point for further understanding the role of ANKs in the virulence mechanisms of plasmodiophorids.

DISCUSSION

This study identified the primary protein folds in gall-forming pathogens secretome supporting the idea that pathogen secretome are often dominated by expansion of specific folds which have been adopted and diversified over the course of evolution. Characterizing these primary effector folds in understudied plasmodiophorids like *P. brassicae* and *S. subterranea* would offer valuable insights into their virulence strategies, which remain largely enigmatic [20,50]. Here we found that the ankyrin proteins, which are significantly expanded in gall-forming plasmodiophorids but less so in the related species *P. betae*, may be key to their ability to manipulate host immune responses and promote gall formation (**Fig. 2**, **Fig. 6**). This finding aligns with previous research indicating the importance of repeat-containing proteins in the virulence of plant pathogens [51]. Example of that are *Phytophthora* spp. effectors containing tandem repeats of the "(L)WY" motif, whose modularity and elaborate mimicry of a host phosphatase helps to promote infection [52]. ANK motifs are well-known for mediating protein-protein interactions [53], and have been identified as type IV effectors in the intracellular human pathogens *Legionella pneumophila* and *Coxiella burnetii* [54].

Thus, we can hypothesize that the highly polymorphic surface residues of plasmodiophorid ANK motifs, coupled with the variable frequency of occurrence in member proteins, results in different binding surfaces for the host target proteins.

This study also identified conserved protein folds across multiple kingdoms, particularly the hydrolase, carboxypeptidase, and aspartyl protease folds (**Table S3**). These folds appear to play a fundamental role in the virulence strategies of these pathogens, likely due to their ability to perform essential biochemical functions that facilitate infection [55–57]. The conservation of these folds across diverse species highlights their evolutionary significance and suggests that they may represent mechanisms of host manipulation that have been retained through speciation. We also identified a nucleoside hydrolase like fold in evolutionary distant gall-forming pathogens which has homology to bacterial effector HopQ1 (**Fig. 3**). Nucleoside hydrolases are involved in the purine salvage pathway in various pathogens [58], but similar to HopQ1's mode of action, these gall-forming biotrophs might also be targeting 14-3-3 proteins, which are implicated in hormonal signaling [59]. Although HopQ1 is widely conserved within the *Pseudomonas* species complex, our FoldSeek-mediated search identified hits in *Pseudomonas savastanoi* isolates, some of which are known to form galls on woody plants [60,61]. It remains to be seen if some of the HopQ1 homologs have been specifically adapted in these bacteria to support a particular lifestyle.

Here we also provided further evidence for the divergent evolution model of effector evolution, which describes how members of the same effector family can exhibit extreme sequence dissimilarity over a long period while retaining the core fold intact [7]. We show that this evolution mechanism occurs in both fungi and oomycetes and often involves the conservation of cysteine or hydrophobic residues to maintain the original fold (**Fig. 4**, **Fig. 5**). Given that sequence dissimilarity between homologs can be extreme, as exemplified by the Mig1 family in *U. maydis*, it would become common practice among researchers to incorporate structural knowledge into sequence searches to accurately gauge the diversity of the effector families (**Fig. 4**).

Overall, our study underscores the power of structural genomics and machine learning tools like AlphaFold2 in uncovering the complexities of pathogen effector repertoires. The findings presented here open new avenues for research into the evolution of virulence strategies in phytopathogens and highlight the potential for these insights to inform the development of novel approaches to plant disease management. As we continue to expand our understanding of effector biology, particularly in under-studied pathogens, it will be crucial to integrate these structural

insights with functional studies to fully elucidate the roles of these proteins in host-pathogen interactions.

MATERIALS AND METHODS

Secretome prediction

The proteome of *P. brassicae* was derived from our recent study generating the first complete genome of the clubroot pathogen [62]. We are thankful to Prof. Anne Legrève for providing the updated annotation of the *P. betae* genome and proteome previously published by them [63,64]. The rest of the proteome for *S. subterranea* [65], *A. candida* [66], *S. endobioticum* [47], *U. maydis* [67] and *T. deformans* [35] were downloaded from Uniprot database. SignalP 6 was utilized to identify sequences with predicted signal peptides, which were subsequently removed. DeepTMHMM was run through pybiolib package. InterProScan 5.61-93.0 was used to confirm the presence of known domains using the Pfam, Gene3D, and SUPERFAMILY databases.

Structure prediction

A total of 3615 mature protein sequences were selected to be modeled using AlphaFold 2, but 40 predictions repeatedly failed at the MSA construction step, resulting in 3575 structures. To expedite the process, ParaFold 2.0 [68] was used, which employs AlphaFold 2.3.1 internally but distributes the CPU and GPU tasks to facilitate parallelization. ‘Valeria’ compute cluster (https://valeria.science/accueil) of Université Laval was used for the structure prediction. The full database was used to construct the MSA (multiple sequence alignment), and models were predicted in 'monomer' mode, resulting in five PDB structures sorted by pLDDT scores. The Rank_0 PDB was used for subsequent studies. Finally, 2000 models with pLDDT scores over 65 were selected for downstream analysis.

Similarity search, clustering, and network plots

TM-Align was used to perform an all-versus-all structural comparison of 2000 models, and those comparisons with a normalized TM-score above 0.5 were considered significant. All-versus-all sequence comparison was performed using BlastP [69] with an E-value < 10^-4 and bidirectional coverage of at least 50%. Structure and sequence similarity data, represented by three columns with the first two as target and source IDs, and the third one being TM-score/E-value, were clustered using the Markov clustering with an inflation value of two. For sequence clustering, E-

values were loaded following the recommendation of the MCL workflow [mcxload -abc seq.abc --stream-mirror --stream-neg-log10 -stream-tf 'ceil(200)']. Custom Python scripts were written to find the sequence-related subclusters belonging to the same structural cluster and to count the occurrence of cluster members (https://github.com/Edelab). Plots were generated using Chiplot (https://www.chiplot.online/) and ggplot2 [70].

Sequence Alignment

Pairwise alignment of protein sequences was done by EMBOSS Needle [71]. Clustal Omega was used to generate multiple protein sequence alignment [72]. Kalign 3.4.0 [73] was used to generate alignment of extremely divergent Mig1 cluster.

Selection pressure analysis and structure visualization

Coding sequences (CDS) were obtained from the Ensembl Fungi/ENA database. Multiple nucleotide sequences were aligned using the Kc-Align codon-aware aligner [74]. Positions with more than 50% gaps were removed using the Clipkit [75] online tool in 'gappy' mode. The trimmed alignments were manually analyzed using Geneious (http://www.geneious.com/) for correct codon alignment. The resulting alignment was uploaded to the Datamonkey server (http://www.datamonkey.org/), which hosts the HyPhy package [76]. All the branches were used as input for FEL [77] to identify sites under purifying selection ($p < 0.01$). The ESPript 3.0 web server was used to generate multiple sequence alignments [78]. Multiple structures were aligned using mTm-Align [79]. PyMOL 3.0.4 [80] was used to visualize the PDB files and color conserved sites or disulfide bridges.

Expression analysis

Datasets for the RNA-Seq reads obtained at 16- and 26 dpi during P. brassicae infection were downloaded from the EBI server (accession number PRJEB12261). The reads from the infected samples were mapped to the *A. thaliana* genome TAIR10 (Genbank accession number GCF_000001735.4) using HiSAT2 [81] to remove host contaminant sequences. To use Salmon [82], an index file was created by concatenating the *P. brassicae* genome and CDS sequences. The remaining RNA-Seq reads were quasi-mapped to the index using Salmon 1.10.0 to generate normalized transcripts per million (TPM) counts for all 10,521 genes. TPM values from three

replicates were averaged. Pre-processed gene expression data for *U. maydis* was publicly available [83].

Motif scanning

MEME 5.5.5 was used to scan the list of amino acid sequences in '-anr' mode (E < 0.1) to discover motifs of any length and frequency. MAST 5.5.5 was used to protein sequences with MEME motifs. Sequence profiles of *P. brassicae* and *S. subterranea* ANK motifs were aligned using Tomtom 5.5.5 [84].

Structural homology search

FoldSeek (https://search.foldseek.com/search) was used to search for structural homology against Uniprot50, Swiss-Prot, the AlphaFold proteome, and PDB. The AFDB cluster database (https://cluster.foldseek.com/) was searched to find cluster members in the AlphaFold database.

In-Silico protein-protein interaction prediction

ANK proteins were screened for interaction against a list of *A. thaliana* immune genes using AlphaPulldown v1.0.4 [85]. It utilizes AlphaFold Multimer but separates the CPU and GPU jobs and reuses the MSA to reduce compute time. The full AlphaFold 2.3.0 database was used for MSA creation. The resulting models from the AlphaPulldown run were parsed with the supplied singularity image alpha-analysis_jax_0.4.sif to produce the final iPTM+pTM score table. ChimeraX 1.8 [86] was used to visualize the predicted aligned error for residue pairs under 4 Ångströms at the interface between the two chains produced by AlphaFold-Multimer.

ACKNOWLEDGMENTS

We are grateful to the bioinformatics support personnel and infrastructure at IBIS, Université Laval, for their constant assistance throughout this project, and to Prof. Sylvain Raffaele for his help in selecting the representative accessions of 62 orphan effector family clusters in Ascomycota. This work was funded by the Canola Agronomic Research Program, Grant ID 2021.4, Western Grain Research Foundation, Canola Council of Canada, Alberta Canola and Manitoba Canola Growers Association; and Discovery Program, Grant ID RGPIN-2021-02518, Natural Sciences and Engineering Research Council of Canada.

COMPETING INTERESTS

None declared.

AUTHOR CONTRIBUTIONS

S.M., M.A.J. and E.P.L. conceived the research; S.M. conducted the research; S.M., and E.P.L. wrote the manuscript and prepared figures; S.M., M.A.J., J.W and E.P.L. edited the manuscript; E.P.L. supervised the research.

ORCID

Soham Mukhopadhyay: 0000-0002-5279-0396

Muhammad Asim Javed: 0000-0003-1658-3565

Jiaxu Wu : 0000-0002-9165-2076

Edel Pérez-López: 0000-0003-2841-9391

DATA AVAILABILITY

The datasets used in this study can be downloaded from Zenodo: https://doi.org/10.5281/zenodo.11152389 , and the scripts from GitHub: https://github.com/Edelab

REFERENCES

1. Toruño, T. Y., Stergiopoulos, I. & Coaker, G. Plant-Pathogen Effectors: Cellular Probes Interfering with Plant Defenses in Spatial and Temporal Manners. *Annu Rev Phytopathol* **54**, 419–441 (2016).

2. Chen, J., Zhang, X., Rathjen, J. P. & Dodds, P. N. Direct recognition of pathogen effectors by plant NLR immune receptors and downstream signalling. *Essays Biochem* **66**, 471–483 (2022).

3. Martel, A. *et al.* The ETS-ETI cycle: evolutionary processes and metapopulation dynamics driving the diversification of pathogen effectors and host immune factors. *Curr Opin Plant Biol* **62**, 102011 (2021).

4. Liu, L. *et al.* Arms race: diverse effector proteins with conserved motifs. *Plant Signal Behav* (2019).

5. Derbyshire, M. C. & Raffaele, S. Surface frustration re-patterning underlies the structural landscape and evolvability of fungal orphan candidate effectors. *Nat Commun* **14**, 5244 (2023).

6. Seong, K. & Krasileva, K. V. Computational Structural Genomics Unravels Common Folds and Novel Families in the Secretome of Fungal Phytopathogen Magnaporthe oryzae. *Molecular Plant-Microbe Interactions®* **34**, 1267–1280 (2021).

7. Seong, K. & Krasileva, K. V. Prediction of effector protein structures from fungal phytopathogens enables evolutionary analyses. *Nat Microbiol* **8**, 174–187 (2023).

8. Guillen, K. de *et al.* Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi. *PLoS Pathog* **11**, e1005228 (2015).

9. Combier, M., Evangelisti, E., Piron, M.-C., Schornack, S. & Mestre, P. Candidate effector proteins from the oomycetes Plasmopara viticola and Phytophthora parasitica share similar predicted structures and induce cell death in Nicotiana species. *PLoS One* **17**, e0278778 (2022).

10. Win, J. *et al.* Sequence Divergent RXLR Effectors Share a Structural Fold Conserved across Plant Pathogenic Oomycete Species. *PLoS Pathog* **8**, e1002400 (2012).

11. Lazar, N. *et al.* A new family of structurally conserved fungal effectors displays epistatic interactions with plant resistance proteins. *PLoS Pathog* **18**, e1010664 (2022).

12. Cao, Y. *et al.* Structural polymorphisms within a common powdery mildew effector scaffold as a driver of coevolution with cereal immune receptors. *Proceedings of the National Academy of Sciences* **120**, e2307604120 (2023).

13. Yu, D. S. *et al.* The structural repertoire of Fusarium oxysporum f. sp. lycopersici effectors revealed by experimental and computational studies. *Elife* **12**, (2024).

14. Teulet, A. *et al.* A pathogen effector FOLD diversified in symbiotic fungi. *New Phytologist* **239**, 1127–1139 (2023).

15. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

16. Nomura, K. *et al.* Bacterial pathogens deliver water- and solute-permeable channels to plant cells. *Nature* **621**, 586–591 (2023).

17. Xin, X.-F. *et al.* Pseudomonas syringae Effector Avirulence Protein E Localizes to the Host Plasma Membrane and Down-Regulates the Expression of the NONRACE-SPECIFIC DISEASE RESISTANCE1/HARPIN-INDUCED1-LIKE13 Gene Required for Antibacterial Immunity in Arabidopsis. *Plant Physiol* **169**, 793–802 (2015).

18. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. Preprint at https://doi.org/10.1101/2021.10.04.463034 (2022).

19. Homma, F., Huang, J. & van der Hoorn, R. A. L. AlphaFold-Multimer predicts cross-kingdom interactions at the plant-pathogen interface. *Nat Commun* **14**, 6040 (2023).

20. Mukhopadhyay, S., Garvetto, A., Neuhauser, S. & Pérez-López, E. Decoding the Arsenal: Protist Effectors and Their Impact on Photosynthetic Hosts. *Molecular Plant-Microbe Interactions®* **37**, 498–506 (2024).

21. Javed, M. A. *et al.* The clubroot pathogen Plasmodiophora brassicae: A profile update. *Mol Plant Pathol* **24**, 89–106 (2023).

22. Ochoa, J. C. *et al.* Natural variation in Arabidopsis responses to Plasmodiophora brassicae reveals an essential role for Resistance to Plasmodiophora brasssicae 1 (RPB1). *The Plant Journal* **116**, 1421–1440 (2023).

23. González-García, M. & Pérez-López, E. Looking for a Cultured Surrogate for Effectome Studies of the Clubroot Pathogen. *Front Microbiol* **12**, (2021).

24. Decroës, A. *et al.* Rhizomania: Hide and Seek of Polymyxa betae and the Beet Necrotic Yellow Vein Virus with Beta vulgaris. *Molecular Plant-Microbe Interactions®* **35**, 989–1005 (2022).

25. Teufel, F. *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* **40**, 1023–1025 (2022).

26. Hallgren, J. *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. Preprint at https://doi.org/10.1101/2022.04.08.487609 (2022).

27. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

28. van de Vossenberg, B. T. L. H. *et al.* The Synchytrium endobioticum AvrSen1 Triggers a Hypersensitive Response in Sen1 Potatoes While Natural Variants Evade Detection. *Molecular Plant-Microbe Interactions®* **32**, 1536–1546 (2019).

29. Erdős, G., Pajkos, M. & Dosztányi, Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res* **49**, W297–W303 (2021).

30. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* **33**, 2302–2309 (2005).

31. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**, 1575–1584 (2002).

32. Van Dongen, S. Graph Clustering Via a Discrete Uncoupling Process. *SIAM Journal on Matrix Analysis and Applications* **30**, 121–141 (2008).

33. Furzer, O. J. *et al.* An Improved Assembly of the Albugo candida Ac2V Genome Reveals the Expansion of the "CCG" Class of Effectors. *Molecular Plant-Microbe Interactions* **35**, 39–48 (2022).

34. Tanaka, S. *et al.* Neofunctionalization of the secreted Tin2 effector in the fungal pathogen Ustilago maydis. *Nat Microbiol* **4**, 251–257 (2019).

35. Cissé, O. H. *et al*. Genome Sequencing of the Plant Pathogen Taphrina deformans, the Causal Agent of Peach Leaf Curl. *mBio* **4**, 10.1128/mbio.00055–13 (2013).

36. Gao, F. *et al.* Deacetylation of chitin oligomers increases virulence in soil-borne fungal pathogens. *Nature Plants 2019 5:11* **5**, 1167–1176 (2019).

37. Petit, M. & Schneider, A. Chemical analysis of the wall of the yeast form of Taphrina deformans. *Arch Microbiol* **135**, 141–146 (1983).

38. Eichhorn, E., Ploeg, J. R. van der, Kertesz, M. A. & Leisinger, T. Characterization of α-Ketoglutarate-dependent Taurine Dioxygenase from Escherichia coli *. *Journal of Biological Chemistry* **272**, 23031–23036 (1997).

39. Thynne, E. *et al*. An array of Zymoseptoria tritici effectors suppress plant immune responses. Preprint at https://doi.org/10.1101/2024.03.12.584321 (2024).

40. Akum, F. N., Steinbrenner, J., Biedenkopf, D., Imani, J. & Kogel, K.-H. The Piriformospora indica effector PIIN_08944 promotes the mutualistic Sebacinalean symbiosis. *Front Plant Sci* **6**, (2015).

41. Lovelace, A. H. *et al*. Effector Identification in Plant Pathogens. *Phytopathology* **113**, 637–650 (2023).

42. van Kempen, M. *et al*. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* **42**, 243–246 (2024).

43. Barrio-Hernandez, I. *et al*. Clustering predicted structures at the scale of the known protein universe. *Nature* **622**, 637–645 (2023).

44. Li, W., Yadeta, K. A., Elmore, J. M. & Coaker, G. The Pseudomonas syringae Effector HopQ1 Promotes Bacterial Virulence and Interacts with Tomato 14-3-3 Proteins in a Phosphorylation-Dependent Manner1[C][W][OA]. *Plant Physiol* **161**, 2062–2074 (2013).

45. Basse, C. W., Stumpferl, S. & Kahmann, R. Characterization of a Ustilago maydisGene Specifically Induced during the Biotrophic Phase: Evidence for Negative as Well as Positive Regulation. *Mol Cell Biol* **20**, 329–339 (2000).

46. Redkar, A. *et al.* The Arabidopsis WRR4A and WRR4B paralogous NLR proteins both confer recognition of multiple Albugo candida effectors. *New Phytol* **237**, 532–547 (2023).

47. van de Vossenberg, B. T. L. H. *et al.* Comparative genomics of chytrid fungi reveal insights into the obligate biotrophic and pathogenic lifestyle of Synchytrium endobioticum. *Sci Rep* **9**, 8672 (2019).

48. Bailey, T. L. & Gribskov, M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**, 48–54 (1998).

49. Geyer, R., Wee, S., Anderson, S., Yates, J. & Wolf, D. A. BTB/POZ Domain Proteins Are Putative Substrate Adaptors for Cullin 3 Ubiquitin Ligases. *Mol Cell* **12**, 783–790 (2003).

50. Pérez-López, E. *et al.* Identification of Plasmodiophora brassicae effectors — A challenging goal. *Virulence* **9**, 1344–1353 (2018).

51. Mesarich, C. H., Bowen, J. K., Hamiaux, C. & Templeton, M. D. Repeat-containing protein effectors of plant-associated organisms. *Front Plant Sci* **6**, 160681 (2015).

52. Li, H. *et al.* Pathogen protein modularity enables elaborate mimicry of a host phosphatase. *Cell* **186**, 3196-3207.e17 (2023).

53. Li, J., Mahajan, A. & Tsai, M. D. Ankyrin repeat: A unique motif mediating protein-protein interactions. *Biochemistry* **45**, 15168–15178 (2006).

54. Pan, X., Lührmann, A., Satoh, A., Laskowski-Arce, M. A. & Roy, C. R. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science (1979)* **320**, 1651–1654 (2008).

55. Bradley, E. L. *et al.* Secreted Glycoside Hydrolase Proteins as Effectors and Invasion Patterns of Plant-Associated Fungi and Oomycetes. *Front Plant Sci* **13**, 853106 (2022).

56. Liu, K. *et al.* Effector Protein Serine Carboxypeptidase FgSCP Is Essential for Full Virulence in Fusarium graminearum and Is Involved in Modulating Plant Immune Responses. *https://doi.org/10.1094/PHYTO-02-24-0068-R* (2024) doi:10.1094/PHYTO-02-24-0068-R.

57. Wang, C. *et al.* The secreted FolAsp aspartic protease facilitates the virulence of Fusarium oxysporum f. sp. lycopersici. *Front Microbiol* **14**, 1103418 (2023).

58. Hofer, A. Targeting the nucleotide metabolism of Trypanosoma brucei and other trypanosomatids. *FEMS Microbiol Rev* **47**, 1–20 (2023).

59. Camoni, L., Visconti, S., Aducci, P. & Marra, M. 14-3-3 Proteins in Plant Hormone Signaling: Doing Several Things at Once. *Front Plant Sci* **9**, (2018).

60. Harmon, C. L. *et al.* Bacterial gall of Loropetalum chinense caused by Pseudomonas amygdali pv. loropetali pv. nov. *Plant Dis* **102**, 799–806 (2018).

61. Ramos, C., Matas, I. M., Bardaji, L., Aragón, I. M. & Murillo, J. Pseudomonas savastanoi pv. savastanoi: some like it knot. *Mol Plant Pathol* **13**, 998–1009 (2012).

62. Javed, M. A., Mukhopadhyay, S., Normandeau, E., Brochu, A.-S. & Pérez-López, E. Telomere-to-telomere Genome Assembly of the Clubroot Pathogen Plasmodiophora Brassicae. *Genome Biol Evol* **16**, evae122 (2024).

63. Decroës, A., Calusinska, M., Delfosse, P., Bragard, C. & Legrève, A. First Draft Genome Sequence of a Polymyxa Genus Member, Polymyxa betae, the Protist Vector of Rhizomania. *Microbiol Resour Announc* **8**, 10.1128/mra.01509–18 (2019).

64. Decroës, A. *et al.* Metagenomics approach for Polymyxa betae genome assembly enables comparative analysis towards deciphering the intracellular parasitic lifestyle of the plasmodiophorids. *Genomics* **114**, 9–22 (2022).

65. Ciaghi, S., Neuhauser, S. & Schwelm, A. Draft Genome Resource for the Potato Powdery Scab Pathogen Spongospora subterranea. *Molecular Plant-Microbe Interactions®* **31**, 1227–1229 (2018).

66. McMullan, M. *et al.* Evidence for suppression of immunity as a driver for genomic introgressions and host range expansion in races of Albugo candida, a generalist parasite. *Elife* **4**, e04550 (2015).

67. Kämper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. *Nature* **444**, 97–101 (2006).

68. Zhong, B. *et al.* ParaFold: Paralleling AlphaFold for Large-Scale Predictions. Preprint at https://doi.org/10.48550/arXiv.2111.06340 (2021).

69. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

70. Wickham, H. ggplot2. (2016) doi:10.1007/978-3-319-24277-4.

71. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276–277 (2000).

72. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Science* **27**, 135–145 (2018).

73. Lassmann, T. Kalign 3: multiple sequence alignment of large datasets. *Bioinformatics* **36**, 1928–1929 (2020).

74. Nicholas Keener. davebx/kc-align: Codon-aware aligner. https://github.com/davebx/kc-align (2020).

75. Steenwyk, J. L., Buida, T. J., Li, Y., Shen, X. X. & Rokas, A. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol* **18**, e3001007 (2020).

76. Kosakovsky Pond, S. L., Frost, S. D. W. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).

77. Kosakovsky Pond, S. L. & Frost, S. D. W. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol Biol Evol* **22**, 1208–1222 (2005).

78. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320–W324 (2014).

79. Dong, R., Peng, Z., Zhang, Y. & Yang, J. mTM-align: an algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics* **34**, 1719–1725 (2018).

80. Schrödinger LLC. *The PyMOL Molecular Graphics System, Version~1.8*. (2015).

81. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).

82. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).

83. Lanver, D. *et al*. The Biotrophic Development of Ustilago maydis Studied by RNA-Seq Analysis. *Plant Cell* **30**, 300–323 (2018).

84. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. *Genome Biol* **8**, R24 (2007).

85. Yu, D., Chojnowski, G., Rosenthal, M. & Kosinski, J. AlphaPulldown—a python package for protein–protein interaction screens using AlphaFold-Multimer. *Bioinformatics* **39**, (2023).

86. Meng, E. C. *et al*. UCSF ChimeraX: Tools for structure building and analysis. *Protein Science* **32**, e4792 (2023).

FIGURE LEGENDS

Fig. 1. Description of the pathogens included in the study, workflow overview and statistics of structure prediction. A, Cladogram of the pathogens used in the study. The schematics represent the disease symptoms on their respective hosts, with the white areas representing galls. The secretome count indicates the number of proteins per species predicted to be secreted, and the

functional annotation shows the percentage of the secretome predicted to contain a known protein domain in the Pfam, SUPERFAMILY, or Gene3D databases. **B**, Flowchart of the workflow used to predict the secretome and the corresponding 3D structures. **C**, Raincloud plot showing the median and density distribution of pLDDT scores of the predicted structures in each pathogen.

Fig. 2. Visualization of dominant protein folds present in each pathogen. A, Network plot of structurally similar secretome clusters with at least 15 members. Not all 255 clusters are shown to reduce complexity. Each node represents a single protein, and an edge between two nodes represents structural similarity (TMScore > 0.5). **B**, Representative structure of the dominant fold in each pathogen. Since Ankyrin repeats are common in both *P. brassicae* and *S. subterranea*, they are represented only once.

Fig. 3. Sequence and structural similarity among HopQ1 homologs in *U. maydis, A. candida* and *P. brassicae.* **A**, Network plot showing the structural similarity between the members of cluster 21. Edges denote structural similarity (TMScore > 0.5). **B**, Pairwise sequence identity between selected HopQ1 structural homologs from plasmodiophorids, oomycetes, and fungi, illustrating sequence dissimilarity between some proteins despite structural homology. **C**, Gene expression values (log2 TPM) of two highly induced *P. brassicae* genes at 16 and 26 dpi. **D**, 3D structure of the mature protein sequences, assuming a HopQ1-like fold.

Fig. 4. Sequence and structural similarity among Mig1 homologs in *Ustilago maydis*. A, Similarity matrix showing the pairwise sequence identity (%) between Mig1 cluster members. **B**, Similarity matrix showing the pairwise structural homology scores (TMScore) between Mig1 cluster members. **C**, Superimposition of two Mig1 homologues, illustrating structural similarity despite extreme sequence divergence. **D**, Differential gene expression patterns of two Mig1 tandem duplicates. **E**, Multiple alignment of protein sequences, highlighting the conservation of cysteine residues (marked in yellow). **F**, Visualization of the conserved cysteine residues forming disulfide bridges.

Fig. 5. SUSS effector families are enriched in common motifs. A, Network plots demonstrating that the two primary effector families in *A. candida* and *S. endobioticum* can only be grouped together when structural data is incorporated into the sequence-based clustering. The plots also

indicate which sequence-based clusters contain the known effectors from these groups. **B**, 'RAYH' and 'CCG' motif patterns identified by MEME scan. **C**, Disulfide bridges in the 'CCG' motif, likely playing a pivotal role in structural maintenance, are highlighted in the virulence factor CCG30. A zoomed-in view of the 'CCG module' shows the four conserved cysteine residues forming disulfide bridge. **D**, The 'RAYH' motif, occupying the central position in the core alpha-helix bundle, is highlighted in six sequence-based subclusters within the AvrSen1-like cluster in *S. endobioticum.*

Fig. 6. Diversity, structural features and host immune targets of ankyrin repeats in Plasmodiophorids. A, Frequency of repeat-containing proteins in *P. brassicae* and *S. subterranea*. **B**, Network plot showing structural homology within *P. brassicae* Ankyrin repeats, also highlighting the Ankyrin repeats with SKP1/BTB/POZ superfamily domains. **C**, Alignment of Ankyrin motifs from *P. brassicae* and *S. subterranea*. **D**, Visualization of conserved hydrophobic residues in a single Ankyrin repeat module. **E**, Number of Ankyrin repeat proteins predicted to target *Arabidopsis* immune proteins. **F**, AlphaFold Multimer predicted complex of MPK3 and PBTT_00818, highlighting the predicted aligned errors of surface contacts under 4 Ångströms.

SUPPLEMENTARY INFORMATION

Table S1. Detailed information about the pathogens investigated in this study and their secretome composition with pLDDT scores, IUPred3 results, and InterproScan identity.

Table S2. Known/orphan effector families searched in gall-forming secretome investigated in this study.

Table S3. Structural clusters generated in this study; ID of the top 26 clusters detected; and frequency of clusters, orphan effectors and known fungal effectors per species studied here.

Table S4. Detailed information of cluster 2, FoldSeek search results, AFDB cluster webtool, and sequence identity matrix with HopQ1.

Table S5. Sequence identity matrix of *Ustilago maydis* Mig1 protein with other members of cluster 30.

Table S6. Expression of *Ustilago maydis mig1* and members of the cluster 30 in axenic, 1, 2, 4, 6, 8, and 12 dpi.

Table S7. Detailed information of the new SUSS effector families identified in this study. Structure based and Sequence based clustering of CCG effectors. Structure based and Sequence based clustering of AvrSen1-like effectors. MEME output of CCG and RAYH motif scan. Selection of one representative member (highest pLDDT) from each sequence-based cluster for selection pressure analysis in CCGs. Selection of one representative member (highest pLDDT) from each sequence-based cluster to examine RAYH motif conservation in AvrSen1-like cluster.

Table S8. Sequence clusters generated in this study.

Table S9. Detailed information of proteins containing repeats. Ankyrin repeat containing proteins, leucine-rich repeats, and Ankyrin repeat containing proteins with SPK1 domains in *P. brassicae* and *S. subterranean*. Expression of *P. brassicae* ankyrin repeat containing proteins and leucine-rich repeats at 16- and 26 dpi in *A. thaliana*. *P. brassicae* and *S. subterranean* ankyrin repeat containing proteins MEME scan.

Table S10. List of *A. thaliana* and *P. brassicae* proteins used to study modeled interactions with AlphaFold-Multimer.

Table S11. Output of the AlphaFold-Multimer predicting interaction between *A. thaliana* immunity hubs and *P. brassicae* Ankyrin repeat containing proteins.

Fig. S1. The structures of *Beta vulgari*'s GTPases, KP6 fold, RALPH-like fold and ToxA fold share structural similarity with candidate effectors found in P. *betae*, *U. maydis* and *T. deformans*,, respectively. The number of clusters where proteins share structural similarities with these folds is provided.

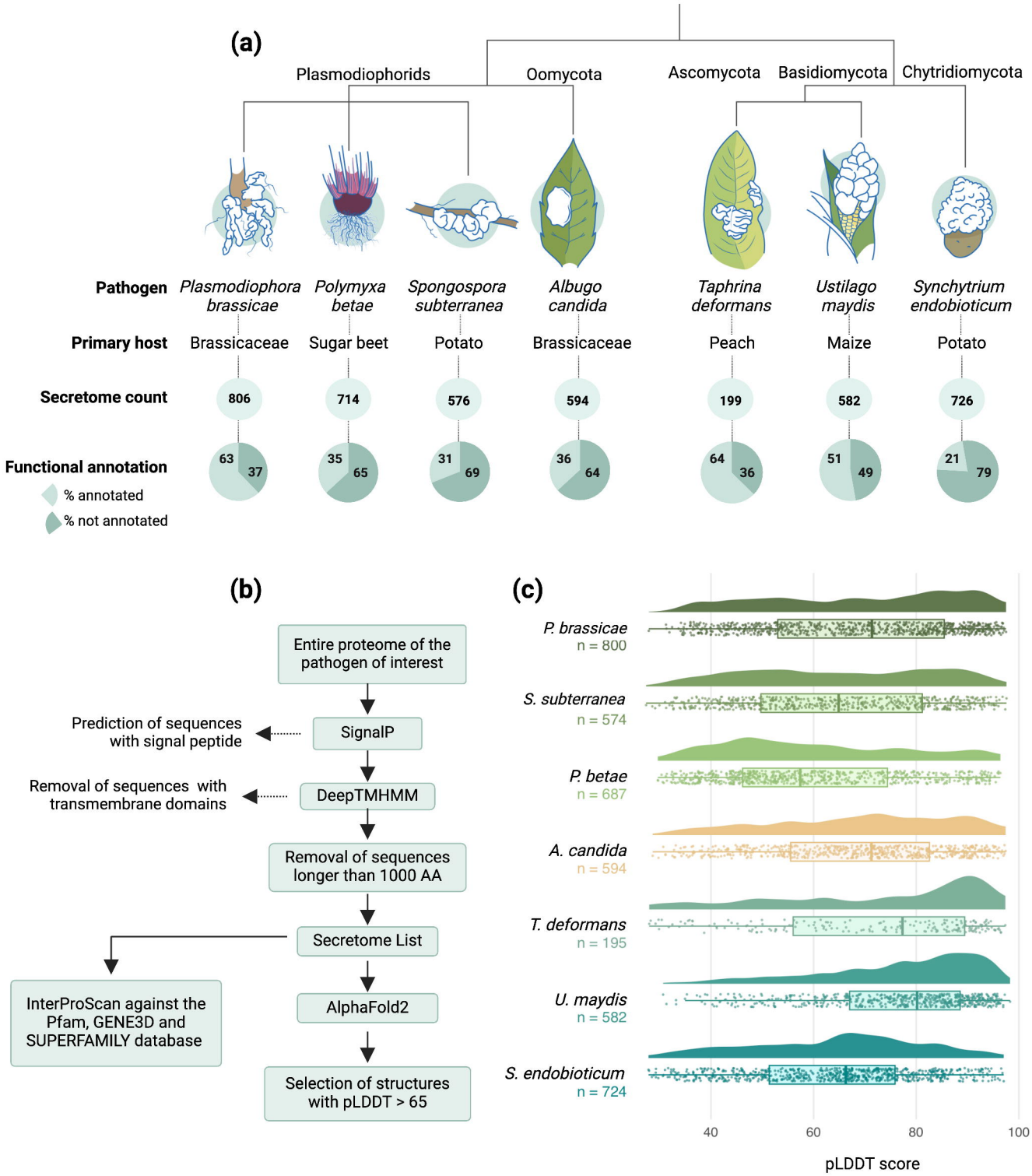
Fig. S2. Location of CCG motifs identified by MEME scan.

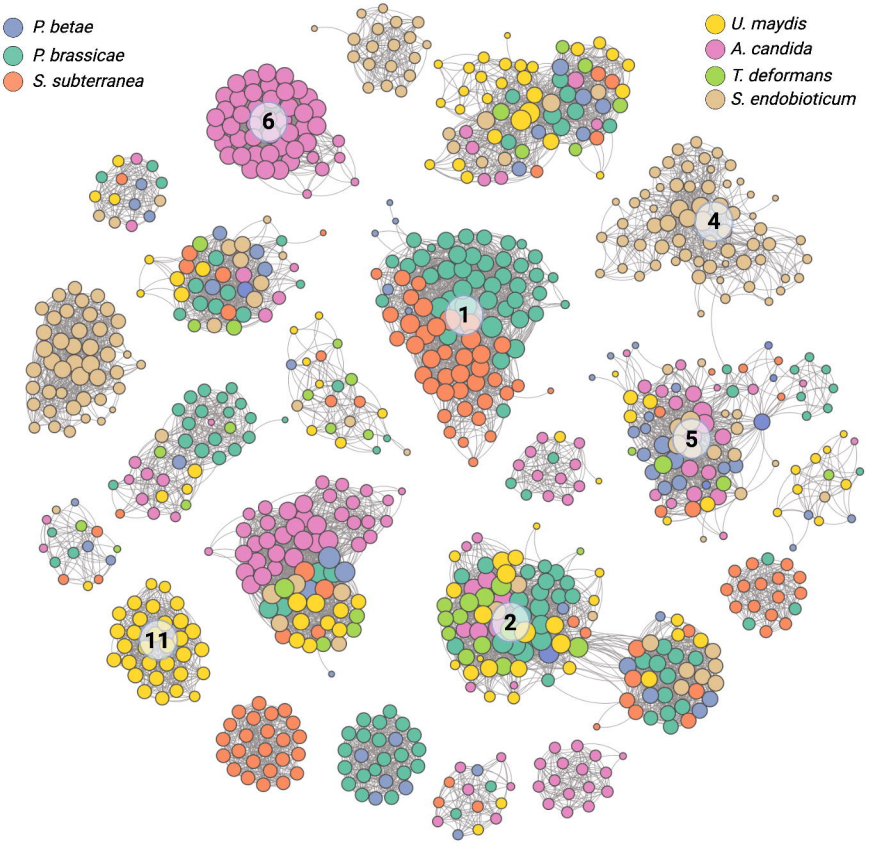
Fig. S3. Location of RAYH motifs identified by MEME scan.

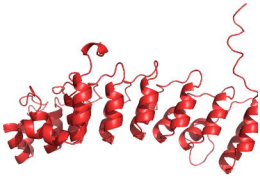
Fig. S4. Analysis evidencing purifying selection in the CCG effector family. Plots have been downloaded from the Datamonkey server, which hosts the FEL package. The indicated codons are numbered according to the codon-aligned nucleotide alignment file, where positions with 50% gaps have been trimmed.

Fig. S5. Multiple alignment of representative members of CCG sequence-based clusters. Conserved residues (>70% similar by biochemical properties in ESPript3.0) are highlighted.

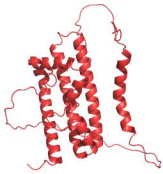
Fig. S6. Multiple alignment of representative members of AvrSen1-like sequence-based clusters. Conserved residues (>70% similar by biochemical properties in ESPript3.0) are highlighted.


(a)
Plasmodiophorids
Oomycota
Ascomycota
Basidiomycota
Chytridiomycota
Pathogen
Plasmodiophora brassicae
Polymyxa betae
Spongospora subterranea
Albugo candida
Taphrina deformans
Ustilago maydis
Synchytrium endobioticum
Primary host
Brassicaceae
Sugar beet
Potato
Brassicaceae
Peach
Maize
Potato
Secretome count
806
714
576
594
199
582
726
Functional annotation
63 37
35 65
31 69
36 64
64 36
51 49
21 79
% annotated
% not annotated
(b)
Entire proteome of the pathogen of interest
SignalP
Prediction of sequences with signal peptide
DeepTMHMM
Removal of sequences with transmembrane domains
Removal of sequences longer than 1000 AA
Secretome List
InterProScan against the Pfam, GENE3D and SUPERFAMILY database
AlphaFold2
Selection of structures with pLDDT > 65
(c)
P. brassicae
n = 800
S. subterranea
n = 574
P. betae
n = 687
A. candida
n = 594
T. deformans
n = 195
U. maydis
n = 582
S. endobioticum
n = 724
40
60
80
100
pLDDT score

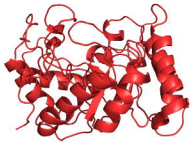




1 PBTT_04511
Ankyrin repeats

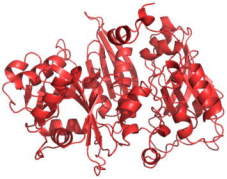


4 A0A507CY85
AvrSen1-like

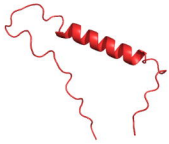


2 R4XM90
CAZYmes

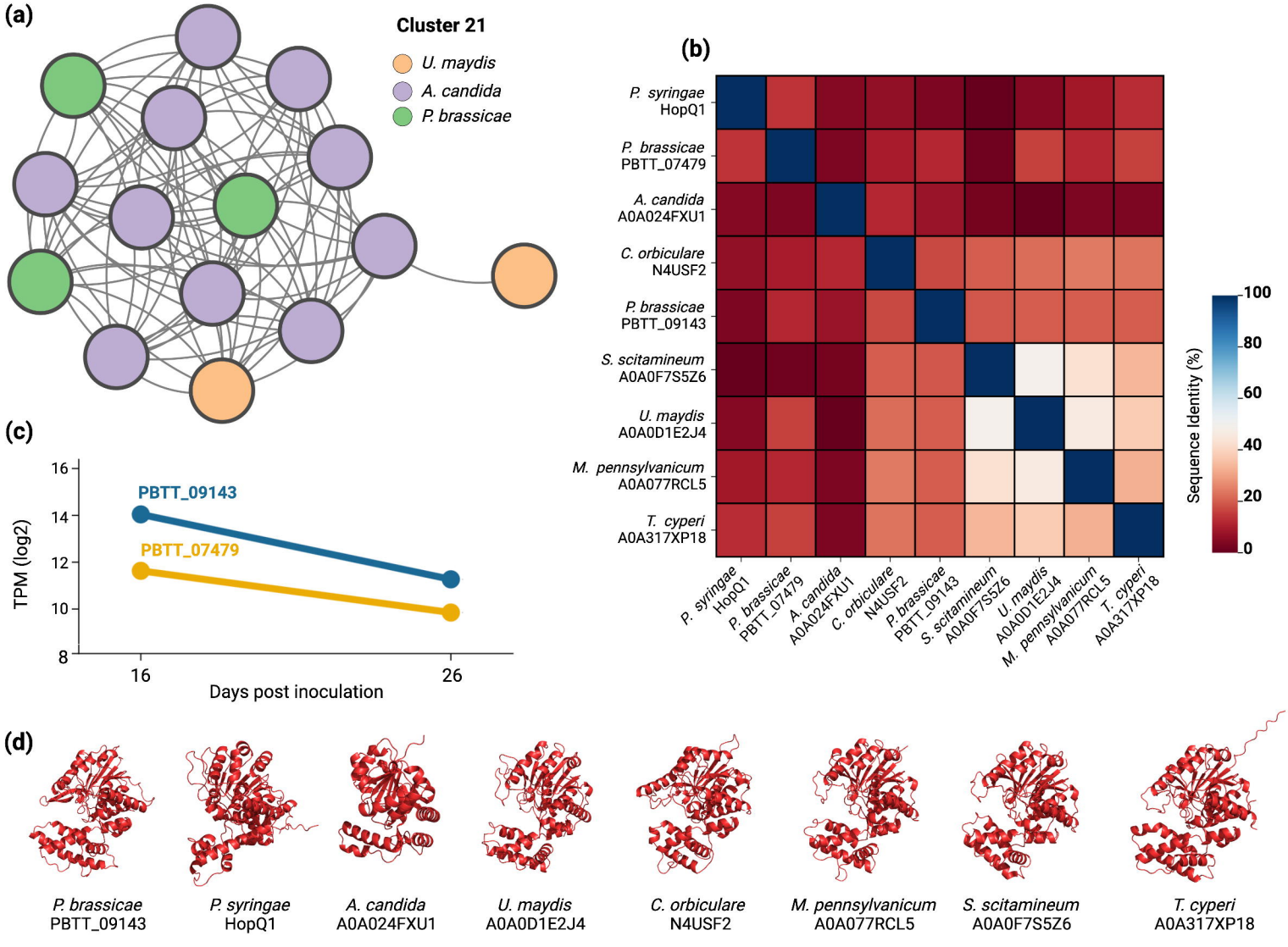
11 UMAG_05302
Tin-2-like

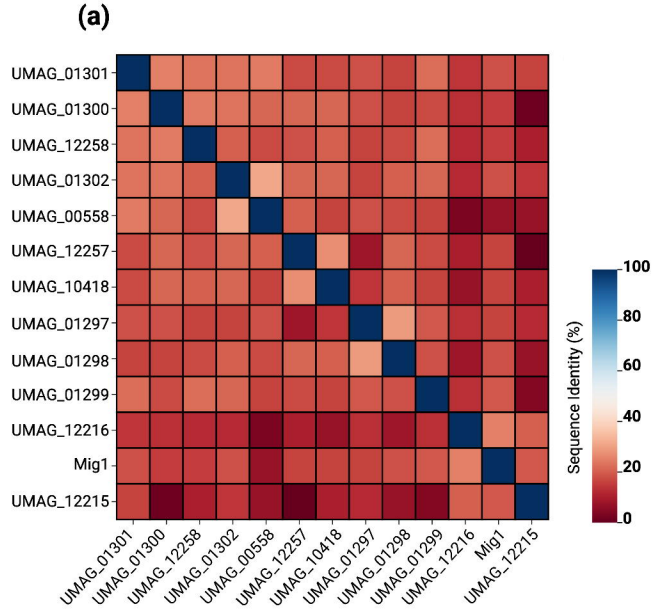


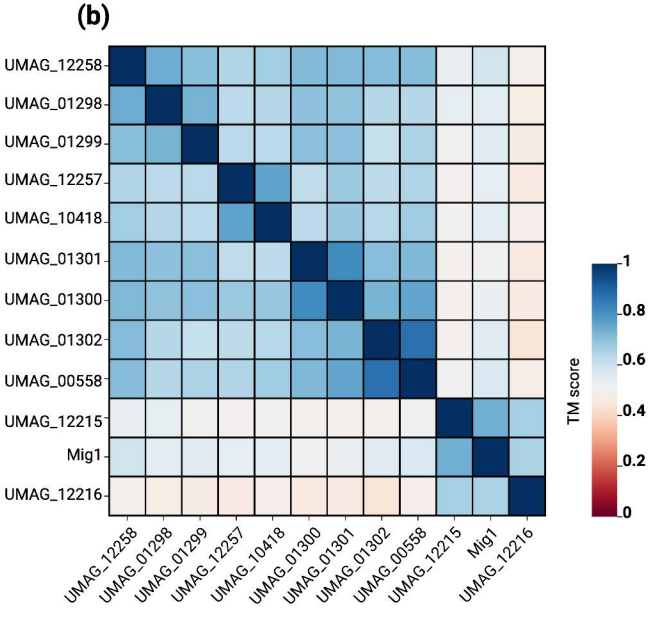
6 A0A024FTM3
CCG family

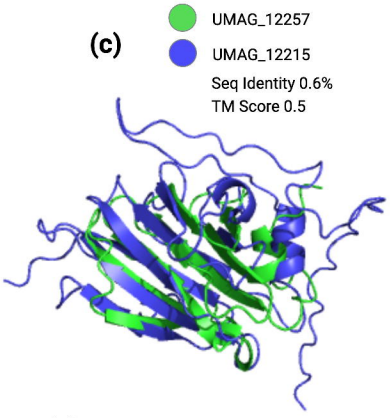


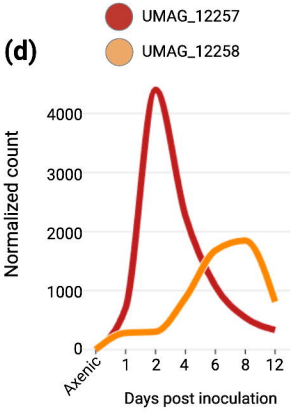
5 JG262
Helical

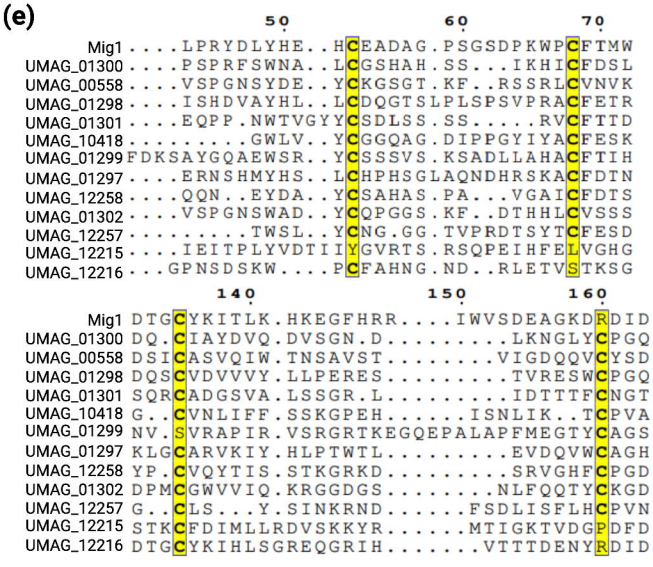

(a)
Cluster 21
U. maydis
A. candida
P. brassicae
(b)
P. syringae HopQ1
P. brassicae PBTT_07479
A. candida A0A024FXU1
C. orbiculare N4USF2
P. brassicae PBTT_09143
S. scitamineum A0A0F7S5Z6
U. maydis A0A0D1E2J4
M. pennsylvanicum A0A077RCL5
T. cyperi A0A317XP18
Sequence Identity (%)
100
80
60
40
20
0
(c)
PBTT_09143
PBTT_07479
TPM (log2)
16
14
12
10
8
16
26
Days post inoculation
(d)
P. brassicae PBTT_09143
P. syringae HopQ1
A. candida A0A024FXU1
U. maydis A0A0D1E2J4
C. orbiculare N4USF2
M. pennsylvanicum A0A077RCL5
S. scitamineum A0A0F7S5Z6
T. cyperi A0A317XP18

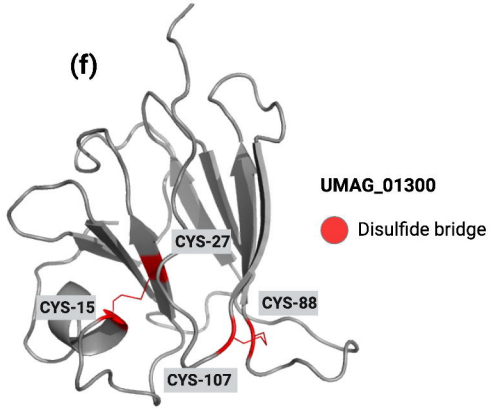

(a)
UMAG_01301
UMAG_01300
UMAG_12258
UMAG_01302
UMAG_00558
UMAG_12257
UMAG_10418
UMAG_01297
UMAG_01298
UMAG_01299
UMAG_12216
Mig1
UMAG_12215
Sequence identity (%)
100
80
60
40
20
0

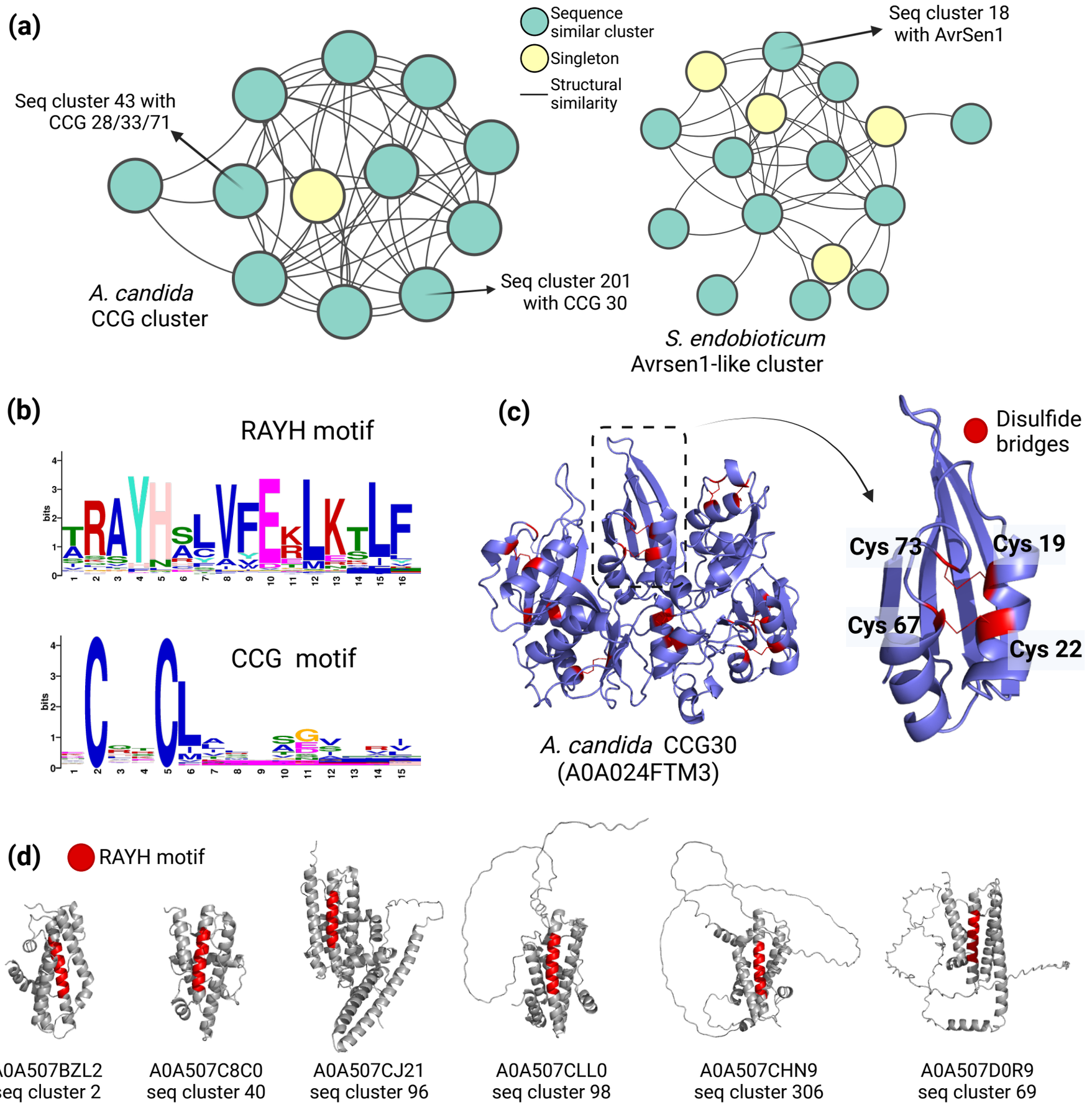

(b)
UMAG_12258
UMAG_01298
UMAG_01299
UMAG_12257
UMAG_10418
UMAG_01301
UMAG_01300
UMAG_01302
UMAG_00558
UMAG_12215
Mig1
UMAG_12216
TM score
1
0.8
0.6
0.4
0.2
0

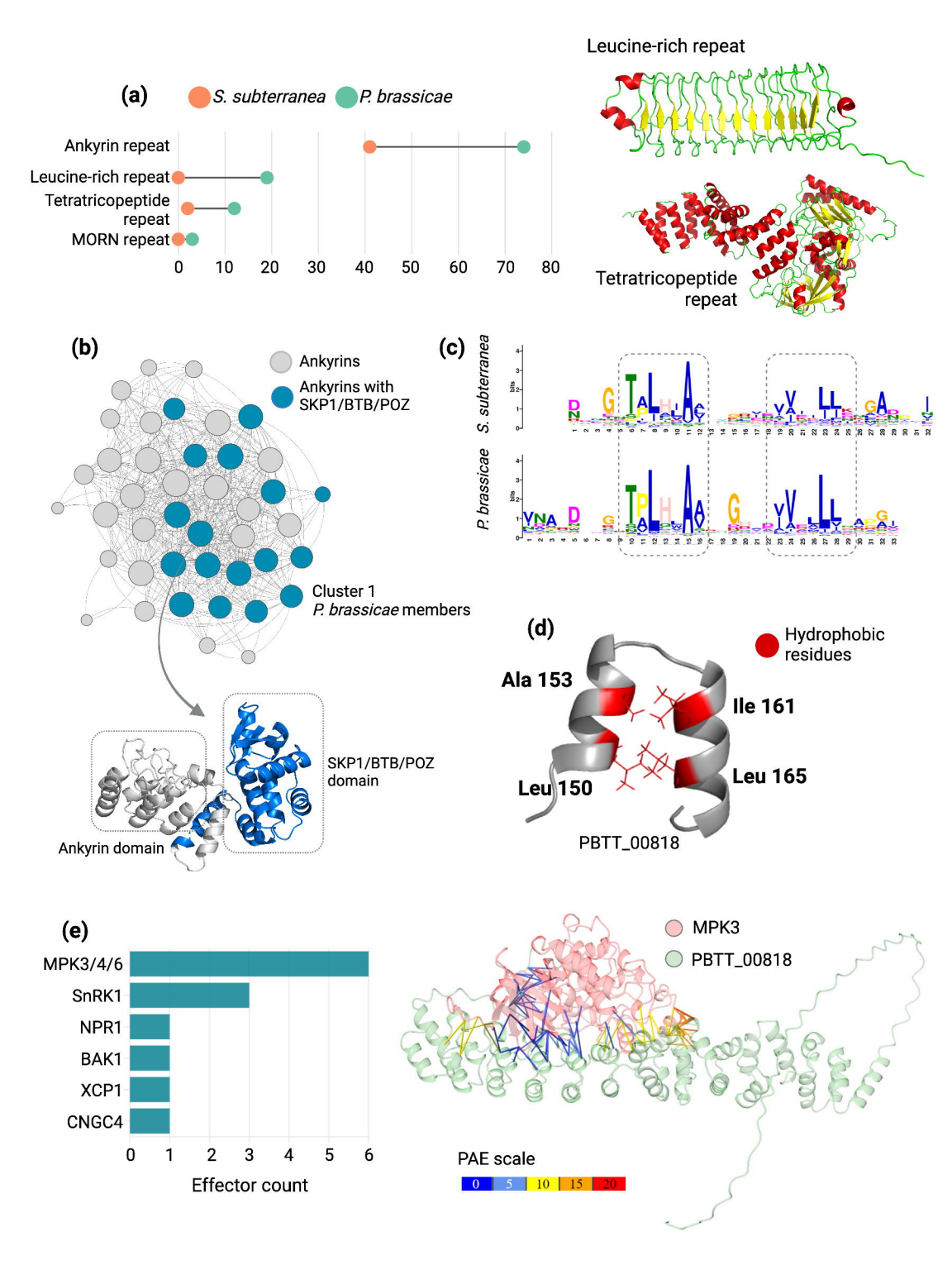

(c)
UMAG_12257
UMAG_12215
Seq Identity 0.6%
TM Score 0.5


(d)
UMAG_12257
UMAG_12258
Normalized count
4000
3000
2000
1000
0
Axenic
1
2
4
6
8
12
Days post inoculation


(e)
50
60
70
Mig1 ....LPRYDLYHE..HCEADAG.PSGSDPKWPCFTMW
UMAG_01300 ....PSPRFSWNA..LCGSHAH.SS...IKHICFDSL
UMAG_00558 ....VSPGNSYDE..YCKGSGT.KF..RSSRLCVNVK
UMAG_01298 ....ISHDVAYHL..LCDQGTSLPLSPSVPRACFETR
UMAG_01301 ....EQPP.NWTVGYYCSDLSS.SS.....RVCFTTD
UMAG_10418 .....GWLV..YCGGQAG.DIPPGYIYACFESK
UMAG_01299 FDKSAYGQAEWSR..YCSSSVS.KSADLLAHACFTIH
UMAG_01297 ....ERNSHMYHS..LCHPHSGLAQNDHRSKACFDTN
UMAG_12258 ....QQN..EYDA..YCSAHAS.PA...VGAICFDTS
UMAG_01302 ....VSPGNSWAD..YCQPGGS.KF..DTHHLCVSSS
UMAG_12257 .....TWSL..YCNG.GG.TVPRDTSYTCFESD
UMAG_12215 ....IEITPLYVDTIIYGVRTS.RSQPEIHFELVGHG
UMAG_12216 ...GPNSDSKW....PCFAHNG.ND..RLETVSTKSG
140
150
160
Mig1 DTGCYKITLK.HKEGFHRR....IWVSDEAGKDRDID
UMAG_01300 DQ.CIAYDVQ.DVSGN.D.....LKNGLYCPGQ
UMAG_00558 DSICASVQIW.TNSAVST.....VIGDQQVCYSD
UMAG_01298 DQSCVDVVVY.LLPERES.....TVRESWCPGQ
UMAG_01301 SQRCADGSVA.LSSGR.L.....IDTTTFCNGT
UMAG_10418 G..CVNLIFF.SSKGPEH.....ISNLIK..TCPVA
UMAG_01299 NV.SVRAPIR.VSRGRTKEGQEPALAPFMEGTYCAGS
UMAG_01297 KLGCARVKIY.HLPTWTL.....EVDQVWCAGH
UMAG_12258 YP.CVQYTIS.STKGRKD.....SRVGHFCPGD
UMAG_01302 DPMCGWVVIQ.KRGGDGS.....NLFQQTYCKGD
UMAG_12257 G..CLS...Y.SINKRND.....FSDLISFLHCPVN
UMAG_12215 STKCFDIMLLRDVSKKYR.....MTIGKTVDGPDFD
UMAG_12216 DTGCYKIHLSGREQGRIH.....VTTTDENYRDID


(f)
UMAG_01300
Disulfide bridge
CYS-27
CYS-15
CYS-88
CYS-107


(a)
Sequence similar cluster
Singleton
Structural similarity
Seq cluster 43 with CCG 28/33/71
Seq cluster 201 with CCG 30
A. candida CCG cluster
Seq cluster 18 with AvrSen1
S. endobioticum Avrsen1-like cluster
(b)
RAYH motif
CCG motif
bits
(c)
Disulfide bridges
Cys 73
Cys 19
Cys 67
Cys 22
A. candida CCG30 (A0A024FTM3)
(d)
RAYH motif
A0A507BZL2 seq cluster 2
A0A507C8C0 seq cluster 40
A0A507CJ21 seq cluster 96
A0A507CLL0 seq cluster 98
A0A507CHN9 seq cluster 306
A0A507D0R9 seq cluster 69


(a)
S. subterranea
P. brassicae
Ankyrin repeat
Leucine-rich repeat
Tetratricopeptide repeat
MORN repeat
0
10
20
30
40
50
60
70
80
Leucine-rich repeat
Tetratricopeptide repeat
(b)
Ankyrins
Ankyrins with SKP1/BTB/POZ
Cluster 1
P. brassicae members
SKP1/BTB/POZ domain
Ankyrin domain
(c)
S. subterranea
P. brassicae
(d)
Hydrophobic residues
Ala 153
Ile 161
Leu 150
Leu 165
PBTT_00818
(e)
MPK3/4/6
SnRK1
NPR1
BAK1
XCP1
CNGC4
0
1
2
3
4
5
6
Effector count
MPK3
PBTT_00818
PAE scale
0
5
10
15
20