

RESEARCH ARTICLE

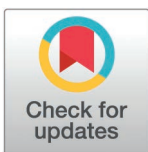
Computational studies reveal structural characterization and novel families of *Puccinia striiformis* f. sp. *tritici* effectors

Raheel Asghar¹*, Nan Wu¹, Noman Ali, Yulei Wang, Mahinur Akkaya¹*

School of Bioengineering, Dalian University of Technology, Dalian, Liaoning, China

* These authors contributed equally to this work.

* msa@dlut.edu.cn



Abstract

Understanding the biological functions of *Puccinia striiformis* f. sp. *tritici* (*Pst*) effectors is fundamental for uncovering the mechanisms of pathogenicity and variability, thereby paving the way for developing durable and effective control strategies for stripe rust. However, due to the lack of an efficient genetic transformation system in *Pst*, progress in effector function studies has been slow. Here, we modeled the structures of 15,201 effectors from twelve *Pst* races or isolates, a *Puccinia striiformis* isolate, and one *Puccinia striiformis* f. sp. *hordei* isolate using AlphaFold2. Of these, 8,102 folds were successfully predicted, and we performed sequence- and structure-based annotations of these effectors. These effectors were classified into 410 structure clusters and 1,005 sequence clusters. Sequence lengths varied widely, with a concentration between 101–250 amino acids, and motif analysis revealed that 47% and 5.81% of the predicted effectors contain known effector motifs [Y/F/W]xC and RxLR, respectively highlighting the structural conservation across a substantial portion of the effectors. Subcellular localization predictions indicated a predominant cytoplasmic localization, with notable chloroplast and nuclear presence. Structure-guided analysis significantly enhances effector prediction efficiency as demonstrated by the 75% among 8,102 have structural annotation. The clustering and annotation prediction both based on the sequence and structure homologies allowed us to determine the adopted folding or fold families of the effectors. A common feature observed was the formation of structural homologies from different sequences. In our study, one of the comparative structural analyses revealed a new structure family with a core structure of four helices, including Pst27791, PstGSRE4, and PstSIE1, which target key wheat immune pathway proteins, impacting the host immune functions. Further comparative structural analysis showed similarities between *Pst* effectors and effectors from other pathogens, such as AvrSr35, AvrSr50, Zt-KP4-1, and MoHrip2, highlighting a possibility of convergent evolutionary strategies, yet to be supported by further data encompassing on some evolutionarily distant species. Currently, our initial analysis is the most one on *Pst* effectors' sequence, structural and annotation relationships providing a novel foundation to advance our future understanding of *Pst* pathogenicity and evolution.

OPEN ACCESS

Citation: Asghar R, Wu N, Ali N, Wang Y, Akkaya M (2025) Computational studies reveal structural characterization and novel families of *Puccinia striiformis* f. sp. *tritici* effectors. PLoS Comput Biol 21(3): e1012503. <https://doi.org/10.1371/journal.pcbi.1012503>

Editor: Turkan Haliloglu, Bogazici University, TÜRKIYE

Received: September 19, 2024

Accepted: February 24, 2025

Published: March 28, 2025

Copyright: © 2025 Asghar 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: 15,201 predicted effectors fasta file, 8,102 AF2 predicted structures of *Pst* effector candidates, and AF2 predicted structures of identified *Pst* effectors, *Pgt* Avr and *Pst* Avr candidates can be downloaded from Zenodo (<https://doi.org/10.5281/zenodo.12946213>).

Funding: This work was supported by a grant from Dalian University of Technology

(DUT18RC(3)050). 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

Stripe rust, caused by the fungus *Puccinia striiformis* f. sp. *tritici* (*Pst*), is a major threat to wheat crops worldwide. The fungus uses special proteins, called effectors, to bypass the plant's immune defenses and establish infection. To better understand how these effectors work, we used a computational tool, AlphaFold2, to predict the structures of over 15,000 *Pst* effector proteins. Interestingly, some of the effectors resemble proteins found in other plant pathogens, suggesting that different fungi may evolve analogously. Our research offers new insights into the combatting strategies of *Pst* and could lead to new methods for protecting wheat from stripe rust.

Introduction

Pathogens have evolved in different ways to evade multiple host defense mechanisms and subvert cellular signaling pathways to facilitate infestation, expansion, and colonization. One tactic is to secrete effector proteins into the host in a spatio-temporally controlled manner. These effectors perform various roles in the apoplast or within the plant cell, such as enhancing pathogen entry, undermining the plant immune system, and altering metabolism [1]. Understanding how effector proteins function in their host is crucial for comprehending the interactions between plants and pathogens [2]. However, fungal effector proteins often lack known functional domains, making them hard to identify. It is challenging to predict their roles based on sequence alone since these proteins have very diverse sequences; rapidly evolving or recently emerged, exhibiting a wide range of variations. This diversity and lack of similarity to known proteins make it difficult to pinpoint potential effector candidates and understand their biological functions.

Although the sequence similarity of effectors is low, it is found that effectors have relatively conservative structures and form structural families [3–5]. The WY domain, a common structural motif in oomycete RXLR effectors, features a conserved α -helical fold stabilized by a hydrophobic core, typically containing Trp (W) and Tyr (Y), as demonstrated by the structural elucidation of two sequence-unrelated *Phytophthora* effectors, *Phytophthora capsici* AVR3a11 and *Phytophthora infestans* PexRD2 [6]. Recently, variants known as LWY domains were noticed, e.g., the *Phytophthora* effector PSR2 has a WY domain and other six variants of WY domains [7]. The ToxA-like structural family encompasses effector proteins such as AvrL567-A and AvrL567-D from *Melampsora lini* [8,9] and Avr2 (SIX3) and SIX8 from *Fusarium oxysporum* [10,11], characterized by their structural similarity to ToxA from *Pyrenophora tritici-repentis* [12]. MAX (*Magnaporthe* AVR) and ToxB-like effectors, featuring a typical six-stranded β -sandwich fold, are crucial for the virulence of *Magnaporthe oryzae* (e.g., AvrPiz-t, AVR1-CO39, AVR-Pia) [13–15] and *Pyrenophora tritici-repentis* (e.g., ToxB) [16], despite their sequence divergence. RALPH (RNase-like proteins associated with haustoria) effectors (e.g., BEC1054) [17], characterized by their RNase-like structure and found predominantly in *Blumeria* fungal species, constitute a notable portion of predicted effectors, showcasing a distinct evolutionary expansion within powdery mildews despite highly divergent sequences [18]. AvrLm4-7 [19] and AvrLm5-9 from *Leptosphaeria maculans* and Ecp11-1 from *Cladosporium fulvum* (now *Fulvia fulva*), despite showing low sequence identity, share the *Leptosphaeria* Avirulence and Suppressing (LARS) fold with candidate effectors predicted in at least 13 different fungi [20]. The *Fusarium oxysporum* f. sp. *lycopersici* (*Fol*) dual-domain (FOLD) effectors, exemplified by Avr1 (SIX4) and Avr3 (SIX1), represent a newly identified structural class of effectors with two distinct domains [11]. However, due to the few effectors

uncovered by structural biology, the counts are approximately 70 for bacteria, 20 for oomycetes, and 80 for fungi (S1 Table), the structural family of effector proteins found is limited. In recent years, based on TrRosetta, AlphaFold2 and other AI tools to predict protein structure and carry out structural classification, it has been found that there are effectors with low sequence similarity among pathogens but whose structural similarity can be classified into the above mentioned known and novel effector structure family [21–27].

Stripe (yellow) rust, caused by *Puccinia striiformis* f. sp. *tritici* (*Pst*), is a serious fungal disease affecting wheat production areas worldwide, posing a significant threat to global food security [28]. Introgressive hybridization breeding for yellow rust resistance (*Yr*) genes is the most effective, environmentally sustainable, and cost-effective strategy to control stripe rust disease [29]. However, with the rapid evolution of new races overcoming specific resistance genes and emerging *Pst* virulence, wheat varieties often lose their resistance in a short period [28]. The rapid variation of *Pst* virulence may be related to its rich effectors and the variability of their subcellular locations within the wheat cell. Genome sequences of multiple *Pst* races or isolates have been analyzed, predicting about 1,000 to 2,000 secretome or effectors for each race [30]. Despite this, since 2011, only about 50 *Pst* effectors (S2 Table) have been identified experimentally [31,32]. Few effectors have been analyzed for their function because of the obligate biotrophic nature of the pathogen requiring a plant host and a lack of an efficient, reliable, and stable transformation system of the urediniospores, making it challenging to study the mechanism of each effector through genetic methods. Nevertheless, the structural analysis assists in determining the function, i.e., Pst_13661 is the only effector protein with a determined structure enabling a combatting strategy for *Pst* [33,34]. The effector that can be specifically recognized by the host nucleotide-binding leucine-rich repeat receptor (NLR) and cause an immune response is the protein encoded by the avirulence gene (*Avr*). The failure of wheat varieties carrying NLR-type *Yr* genes to resist *Pst* may be related to *Avr* mutations, which render the *Yr* protein unrecognizable and unable to trigger immunity. Currently, no *Avr* of *Pst* has been identified. However, five *Avr* genes have been identified in *Puccinia graminis* f. sp. *tritici* (*Pgt*), a close relative of *Pst*, the causal agent of stem rust [35]. These *Avr* genes, *AvrSr50* [36,37], *AvrSr35* [38–40], *AvrSr22* [41], *AvrSr13* [41], and *AvrSr27* [42,43], have been confirmed to be recognized by their corresponding wheat NLRs: *Sr50*, *Sr35*, *Sr22*, *Sr13*, and *Sr27*. Recently, it has been reported that the structure of pathogen-secreted proteins can be predicted using AlphaFold2 or other tools and annotated by protein structure databases such as PDB, CATH, and SCOP [23,44,45]. One reason for the slow progress in identifying *Pst* effectors is that predicted effector sequences often lack functional annotations and domain information on protein annotation websites, making it difficult to understand their molecular mechanisms. So far, no study has been reported focusing on large-scale structure predictions or structural annotations of *Pst* effectors.

In this study, by analyzing twelve *Pst* races and isolates, one *Puccinia striiformis* f. sp. *hordei* isolate, and one *Puccinia striiformis* isolate, which consists of 21 protein sets in total, we collected 357,396 proteins and predicted 15,201 effector proteins based on their sequences. Of these, 8,102 had high-confidence predicted structural folds, resulting in the identification of 410 structure clusters and 1,005 sequence clusters. Among the 8,102 effectors, 20.9% have sequence annotations, 75.4% have structure annotations, 6.7% are sequence-related, and 44.2% are structurally related to identified *Pst* effectors. In addition to conventional methods of effector characterization, this structural annotation approach can significantly enhance the efficiency and comprehensiveness of effector analysis. Remarkably, we discovered *AvrSr35*-like, *AvrSr50*-like, *Zt-KP4-1*-like, and *MoHrip2*-like *Pst* effector candidates with little or no sequence similarity, yet they exhibited conserved structural features. Understanding the structure and relationships of *Pst* effector proteins enhances our insight into their biological

functions. This knowledge will be crucial in unraveling the pathogenic mechanisms of wheat stripe rust and in developing new control strategies.

Results

Structure prediction of *Pst* effector candidates with AF2

357,396 proteins were collected from 21 proteomes ([S3 Table](#), where the references are presented) within 14 *Puccinia striiformis* races or isolates: 12 from *Puccinia striiformis* f. sp. *tritici* (*Pst*), 1 from *Puccinia striiformis* isolate 11-281, and 1 from *Puccinia striiformis* f. sp. *hordei* 93TX-2. We predicted the proteins with signal peptides as secretomes using SignalP 6.0 [46] and excluded proteins annotated with transmembrane domains in InterPro [47] or predicted to have a glycosyl-phosphatidyl-inositol (GPI)-anchor by NetGPI [48], resulting in the prediction of 27,444 secreted proteins. Further prediction with EffectorP 3.0 was performed [49], which predicts effector proteins by analyzing sequence features including amino acid composition, sequence motif, and other physiochemical properties known to correlate with effector function, then identifies putative effectors based on machine learning model trained on fungal effector datasets. Redundancy removal across all races identified 15,201 effector candidates. Structural predictions for the mature sequence (excluding the signal peptide) of these effector candidates were conducted using AlphaFold2 (AF2) [50,51] ([Fig 1A](#)). We filtered out structures with pTM (predicted Templated Modeling) scores < 0.5 and pLDDT (predicted Local Distance Difference Test) average scores across all residues < 70, retaining high-confidence models. Following AF2 modeling for 15,201 predicted effectors, nearly half could not be reliably modeled across the 14 races or isolates ([Fig 1B](#) and [1C](#)). This may be due to the high specificity of some effectors, which challenges their alignment with sufficient multiple sequence alignment and template structures during AF2 modeling; additionally, the presence of intrinsically disordered regions in some effectors could also result in low pTM or pLDDT scores. This process yielded 8,102 effectors with reliable structural predictions across 14 races or isolates ([S3 Table](#)). We then performed clustering, annotation, and comparative analysis of these 8,102 effectors from both sequence and structural perspectives for further investigation ([Fig 1A](#)). Among the 8,102 effector candidates, sequence lengths ranged from 43 to 939 amino acids, with a concentration of 2,221 effectors within the length 101–250 amino acids, consistent with the typical feature of effector lengths ([Fig 1D](#)).

Clustering of *Pst* effector candidates with sequence and structural comparison

We performed sequence clustering on the mature sequences of 8,102 effector candidates with high-confidence structural predictions using CD-HIT [52] with a threshold of 0.5, it refers to a 50% sequence identity threshold in CD-HIT clustering, resulting in 1,005 sequence clusters ([S4 Table](#)). Additionally, we used Foldseek release (8-ef4e960) [53] easy cluster function for predicted structures of 8,102 effector candidates with a threshold of 0.5 (it refers to a 50% sequence alignment coverage) resulting in 410 structure clusters ([S4 Table](#)). Clusters are ordered according to the number of effectors they contain, with Structure Cluster No. 1 (Struc.C_1) having the most effector candidates. Of these 410 structure clusters, 165 were singletons. 7,929 effectors were classified as shared since they belonged to the structure cluster containing effectors from two or more *Ps* races or isolates, and 173 effectors were classified as race-specific since they belonged to the structure cluster containing effectors only from one *Ps* race or isolate ([Fig 1C](#)). We analyzed the distribution of 8,102 effectors within sequence and structure clusters across 14 races or isolates using cluster heatmaps and principal component analysis (PCA). In general, the quantitative characteristics and classification relationships of

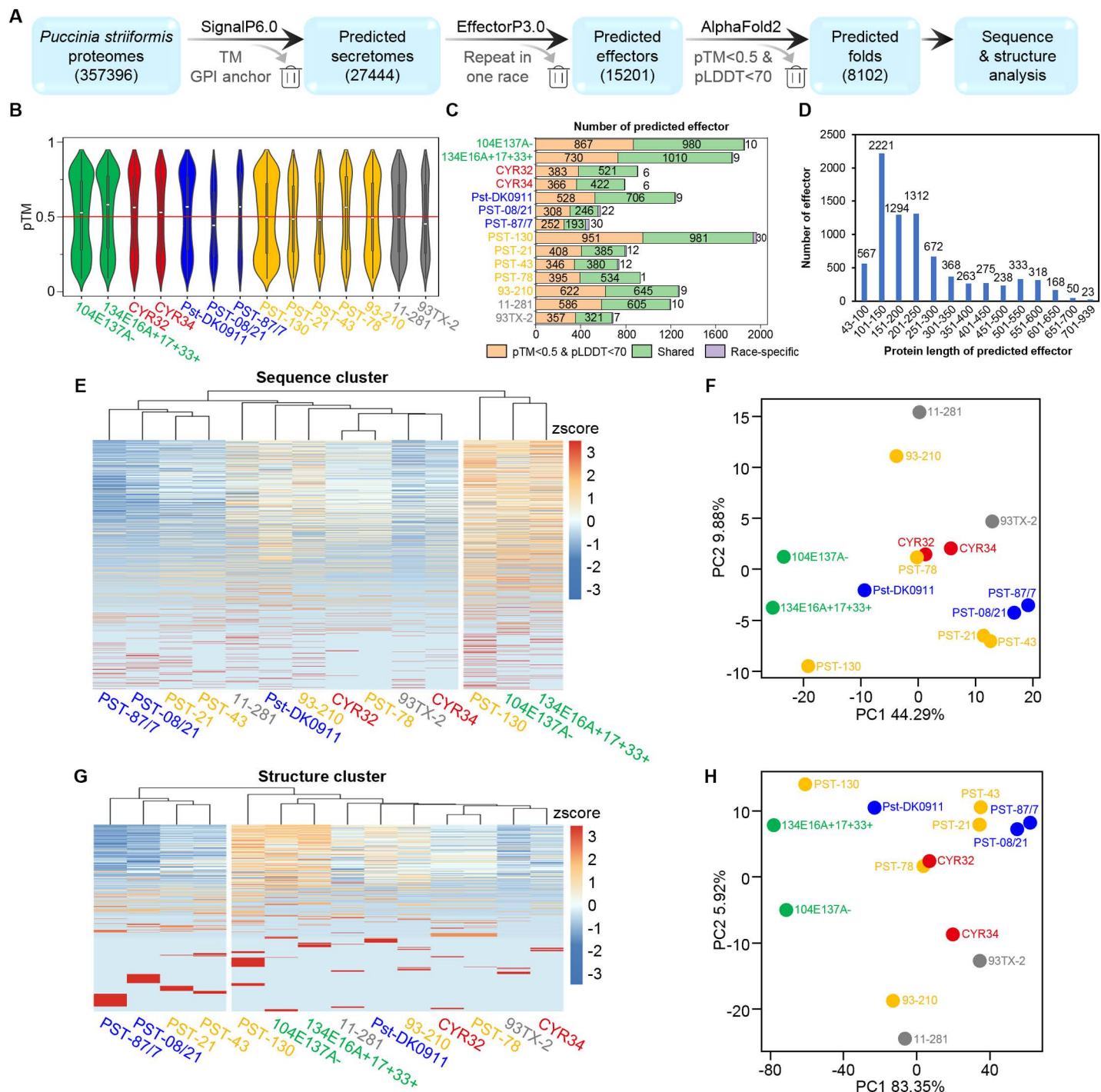


Fig 1. Effector prediction pipeline and statistical analysis of prediction and clustering. (A) Bioinformatics workflow for predicting and analyzing effectors in *Puccinia striiformis* (Ps) proteomes. Blue boxes denote prediction steps with the number of retained proteins in brackets. Analytical tools are above the arrow; filtered steps are below. Out of 15,201 predicted effectors, 8,102 with reliable structures predicted for further investigation. (B) Distribution of pTM scores for evaluating the quality of structure predictions among the 15,201 predicted effectors. The box plot within the violin plot shows the interquartile range (25th to 75th percentiles), with the median represented by a white dot. The whiskers extend to the minimum and maximum values, capped at 1.5 times the interquartile range. The red line represents a pTM value of 0.5; structures with pTM scores below 0.5 were removed. (C) The number of 15,201 predicted effectors was categorized into 14 Ps races or isolates and 'race-specific' otherwise. The effectors are categorized as 'shared' if they belonged to the structure cluster containing effectors from two or more Ps races or isolates and 'race-specific' otherwise. Orange indicates proteins that did not fold well, green denotes 'shared' effectors, and purple denotes 'race-specific' effectors. (D) Sequence length distribution of the 8,102 effectors with reliable structure predictions. (E, F) Cluster heatmap and principal component analysis (PCA) show the number of effectors from

each cluster of 1,005 sequence clusters assigned within 14 *Pst* races or isolates. (G, H) Cluster heatmap and PCA showing the number of effectors from each cluster of 410 structure clusters assigned within 14 *Pst* races or isolates. *Pst* races or isolates are colored based on their region of first discovery: green for Australia, red for China, blue for Denmark and the UK, yellow for the US, and grey for *Puccinia striiformis* isolate 11-281 and *Puccinia striiformis* f. sp. *hordei* isolate 93TX-2. The same color scheme is applied in [Figs 4–6](#).

<https://doi.org/10.1371/journal.pcbi.1012503.g001>

effectors in 14 races or isolates exhibit consistency in both sequence and structure clusters, suggesting that effectors with similar sequences tend to form similar structures. Notably, there is a close correspondence between the sequence clusters and structure clusters of effectors from CYR32 and PST-78, indicating similarity in effector components between these two *Pst* races. Furthermore, the relationships between sequence clusters and structure clusters of effectors from PST-87/7, PST-08/21, PST-21, and PST-43 show similarities. Interestingly, despite having different hosts, CYR34 and 93TX-2 share similarities in terms of both sequence and structure for their effectors ([Fig 1E–1H](#)).

A comprehensive understanding of *Pst* effector candidates' characterization with sequence-based and structural annotation

Cysteine richness is a characteristic feature of effectors. We analyzed the cysteine content in the mature sequences of the 8,102 effectors. Among these, 1,970 effectors contain 6 cysteine residues. Cysteine residues form disulfide bonds that stabilize the effector structure, enabling it to function in harsh environments like the apoplast and resist proteolytic degradation [54,55]. Remarkably, 4 effectors have 30 cysteine residues, 13 of which have 31 cysteine residues ([Fig 2A](#)) with an average length of 458 amino acids. These 17 effectors belong to Struc.C_27 and Sequence Cluster No. 62 (Seq.C_62). They are predicted to be apoplastic effectors by ApoplastP [49,56] but also predict to contain nuclear localization signals using LOCALIZER [57] and WoLF PSORT [58] ([S4 Table](#)). Additionally, many short effectors (<100 amino acids) have a higher cysteine content, such as effectors belonging to Struc.C_27 and Seq.C_62, the average length of mature sequences is 55 amino acids and contain 8 cysteines, with a cysteine content of 14.5% ([S4 Table](#)).

Classical effector motif analysis ([Fig 2B](#) and [S4 Table](#)) revealed that 3,804 effectors (approximately 47% among 8,102 effector candidates) contain the [Y/F/W]xC motif, which is found in the wheat powdery mildew and rust effector candidates, predominantly in the forms of FxC and WxC. Additionally, 896 effectors possess the [L/I]xAR motif, which is found in some effectors of *Magnaporthe oryzae*, with LxAR being the most common. The RxLR motif, a common effector motif from oomycetes and fungi, is present in 471 effectors. Another 223 effectors contain the G[I/F/Y][A/L/S/T]R motif, which is found in some effectors of *Melampsora lini*. A total of 29 effectors have the YxSL[R/K] motif, which is found in oomycetes. We checked motif [R/K]CxxCx12H which is found in *Magnaporthe oryzae* and [R/K]VY[L/I]R, which is found in *Blumeria graminis* f. sp. *hordei*, but not detected. This indicates that different pathogens have their own specific motif characteristics.

We used EffectorP 3.0 [49], ApoplastP [56], LOCALIZER [57], WoLF PSORT [58], and TargetP 2.0 [59] for subcellular localization prediction of the 8,102 effectors ([Fig 2C](#) and [S4 Table](#)). Although the predictions varied across different programs, the overall trend indicated that more effectors were predicted to be cytoplasmic rather than apoplastic, with primary localizations in the chloroplast and nucleus. Notably, 141 effectors' mature sequences were predicted to have signal peptides by TargetP 2.0, but SignalP 6.0 did not detect any signal peptides in these mature effector sequences. Therefore, when predicting subcellular localization using different programs, it is essential to conduct a comprehensive and integrated analysis to ensure the accuracy of the predictions.



Fig 2. Statistical analysis of effector characteristics. (A) Statistics of cysteine count in the mature sequences of effectors. (B) Number of motif-containing effectors. (C) Number of effectors in different subcellular localization predictions. (D) Number of effectors in various protein sequence annotation databases, and statistics of the top 17 hits of effectors annotated within Pfam. (E) Statistics of the top 17 hits of effectors structurally annotated within CATH, PDB, and SCOP.

<https://doi.org/10.1371/journal.pcbi.1012503.g002>

The sequence- and structure-based annotations for the effector candidates were obtained from publicly available databases. We annotated the sequences of effector candidates using various protein databases, including PANTHER [60], Pfam [61], SUPERFAMILY [62], Gene3D [63], Coils [64], ProSiteProfiles and ProSitePatterns [65], CDD [66], MEROPS [67], KEGG [68], FunFam [69], SMART [70], PRINTS [71], CAZy [72], NCBIfam [73], Hamap [74], and PIRSF [75] (Fig 2D and S4 Table). Overall, 1,695 effectors (approximately 21% among 8,102 effector candidates) had sequence annotation information. Specifically, 802 effectors were annotated by PANTHER, 672 by Pfam, 639 by SUPERFAMILY, and 617

by Gene3D. Among all protein databases, Pfam provided the most annotation entries (19k entries, accessed on July 18, 2023). Since Pfam includes the most comprehensive and numerous annotations, we examined the top statistics from Pfam to know sequence annotation of effector candidates in general. The top three Pfam annotations were trehalose-phosphatase, copper/zinc superoxide dismutase, and glycoside hydrolase 131 catalytic N-terminal.

In addition to sequence-based annotation, we utilized the predicted structures of effector candidates using AF2 and compared them with protein structure annotation information from CATH (Protein Structure Classification Database; Class, Architecture, Topology, Homologous superfamily) [63], PDB (Protein Data Bank) [76], and SCOP (Structural Classification of Proteins) [77,78] by Foldseek. This approach provided structural annotation information for 6,110 effectors (approximately 75% among 8,102 effector candidates), significantly more than sequence-based annotations (S4 Table). Specifically, 5,112 effectors were annotated by CATH, 4,556 by PDB, and 5,106 by SCOP. We analyzed the top 20 hits for each database. The most frequently occurring annotations were immunoglobulin in CATH, Hce2 domain-containing protein (or named as *Zymoseptoria tritici* effector Zt-KP4-1) in PDB, and membrane fusion ATPase p97 N-terminal domain in SCOP (Fig 2E). Although the annotation methods of CATH, PDB, and SCOP differ and no identical annotations appear in the top 20 hits of all three databases, certain annotations were common between PDB and SCOP, such as superoxide dismutase, cytochrome c, and trehalose-6-phosphate phosphatase. Superoxide dismutase and trehalose-phosphatase are also the major annotations in sequence annotation. These structural annotations indicate that the effector candidates may adopt similar folds to proteins such as superoxide dismutase, but further analysis would be required to determine their functional roles or family membership.

***Pst* effector candidates reflect progressively differentiating structure**

From the 8,102 effectors, we selected 1,178 representative effectors (marked in grey in S4 Table) from distinct structure and sequence clusters to perform a pair-wise TM-align analysis using Foldseek by filtering out pairs with TM scores < 0.5. This resulted in a structure cluster network graph with 1,015 nodes and 32,119 edges. We present the representative structures of the top ten largest structure clusters (Fig 3A). The largest structure cluster is Struc.C_1 containing 683 effectors. Struc.C_2 to Struc.C_5 contains 438, 326, 323, and 254 effectors, respectively. The top ten structure clusters together account for 36.6% of the effectors, representing the overall characteristics of *Pst* effector candidates. By observing the structure cluster network, we can identify the expansion and variation patterns of effector structures. For example, the structures of Struc.C_10 and Struc.C_1 are quite similar, but Struc.C_10 forms multiple tandem repeats of the Struc.C_1 structure. Similarly, Struc.C_3 appears as a double tandem structure of Struc.C_9. Struc.C_2 shows a dispersed expansion trend, potentially indicating faster structural variation and the gradual formation of new effector clusters like Struc.C_7. The effector structures of Struc.C_4, Struc.C_6, and Struc.C_8 are similar but have gradually diverged structurally. In contrast, Struc.C_5 has a relatively simple structure with fewer associations with other major structure clusters.

Effector structures are widely conserved despite sequence variability

Our analysis of the correspondence between structure and sequence clusters (S5 Table) reveals that effectors within the same sequence cluster are predominantly distributed into a single structure cluster. This supports the general expectation that similar sequences tend to adopt similar structures, as observed in the cluster heatmap analysis (Fig 1E and 1G). However, a single structure cluster can encompass multiple sequence clusters. For instance,

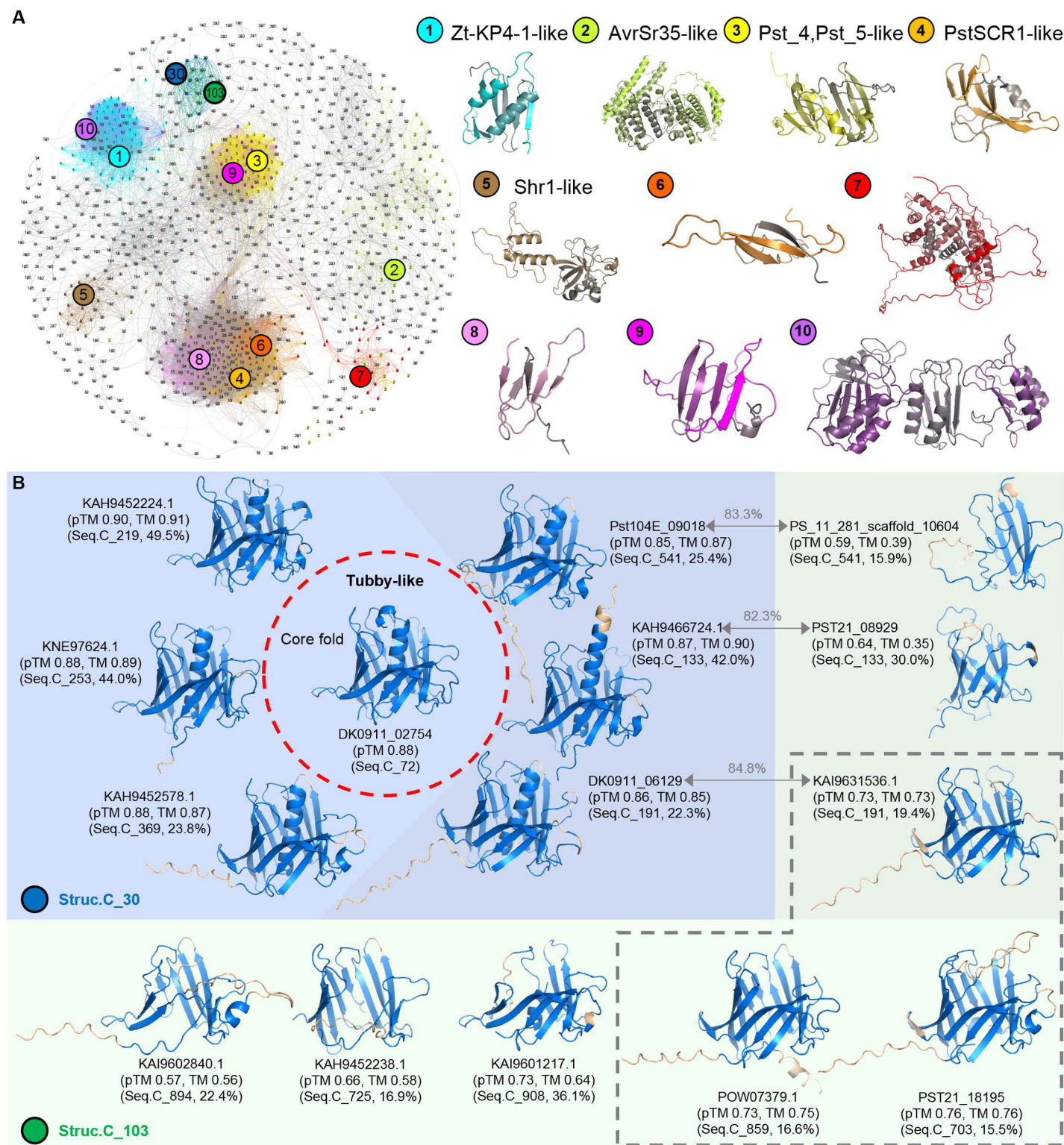


Fig 3. Relationship of *Pst* effector structure clusters. (A) Structural similarity network of representative effectors from each structure cluster's sequence cluster (1,178 proteins marked in grey in [S4 Table](#)). The 10 major structure clusters, along with structure clusters No. 30 (Struc.C_30) and No. 103 (Struc.C_103), are highlighted. Representative structures from the 10 major structure clusters are shown. (B) Structures from Struc.C_30 are shown with a blue background, and structures from Struc.C_103 are shown with a green background. The effector DK0911_02754, highlighted within the red dashed circle, is annotated as Tubby-like based on its sequence, with its predicted structure used as the core fold. Parts of other displayed structures that do not superimpose with the core fold are marked in light orange.

The transparent grey background indicates three pairs of effectors from the same sequence cluster but belonging to Struc.C_30 and Struc.C_103, respectively. The sequence similarity between these pairs is indicated above the connecting arrows. Effectors within the grey dashed box are from Struc.C_103 and lack the central α -helix found in Struc.C_30 members. The pTM values from AF2 modeling, TM-scores (TM) compared with the core fold, sequence cluster numbers (Seq.C), and sequence identity with the core fold sequence are shown below the protein IDs.

<https://doi.org/10.1371/journal.pcbi.1012503.g003>

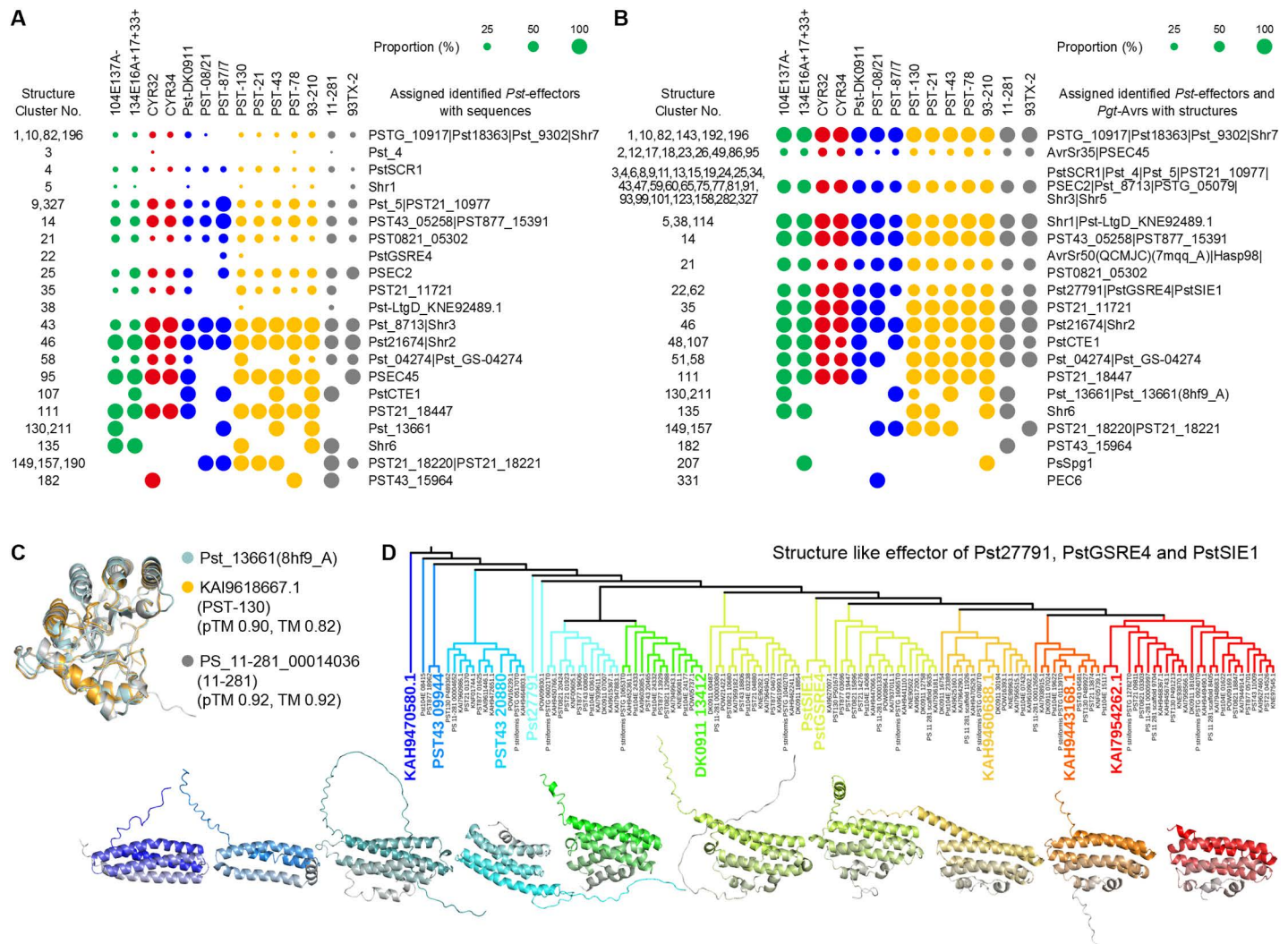


Fig 4. Sequence-based, structural and phylogenetic analysis between identified *Pst* effectors and *Pst* effector candidates. (A) The proportion of identified *Pst* effectors assigned to effector candidates from 14 *Ps* races or isolates within structure clusters, based on sequence comparison, shown with circles of varying sizes. (B) Proportion of identified *Pst* effectors with predicted structures and PDB structures of Pst_13661 (8HF9_A), AvrSr35 (7XX2_B, 7XC2_D, 7XDS_A, 7XDS_B, 7XE0_B, 7XVG_B), and AvrSr50(QCMJC) (7MQQ_A) assigned to predicted structure of effector candidates from 14 *Ps* races or isolates within structure clusters, shown with circles of varying sizes. (C) Pst_13661 (8HF9_A) structure and Pst_13661-like effector predicted structures. The pTM values from AF2 modeling and TM-scores (TM) compared with the Pst_13661 (8HF9_A) structure are indicated below the protein IDs. (D) Structural phylogenetic tree analysis of predicted structures of identified *Pst* effectors Pst27791, PstGSRE4, PstSIE1 and effector candidates from Struc.C_22 and Struc.C_62. Representative structures of groups are shown in the same color as their groups in the phylogenetic tree, with structures displayed from N-terminus (colored) to C-terminus (grey).

<https://doi.org/10.1371/journal.pcbi.1012503.g004>

Struc.C_2, Struc.C_1, and Struc.C_3 contain 61, 52, and 41 sequence clusters, respectively (S5 Table). This suggests that different effector sequences can also form similar structural conformations. To further quantify the extent of sequence diversity within structure clusters, we calculated pairwise sequence identity distributions for all members within each structure

cluster except singleton cluster using CLUSTALW (S5 Table). In large structure clusters, the average sequence identities of Struc.C_1 to Struc.C_5 are 18.22%, 14.34%, 18.44%, 19.20%, and 31.05%, respectively. These results highlight that while structural similarity is generally conserved, the sequence identity among effectors within the same structural cluster can be quite low. This systematic sequence identity analysis further supports the notion that structure is more conserved than sequence in *Pst* effectors.

Taking Struc.C_30 as an example, it includes 7 sequence clusters, with protein sequence similarities between clusters being less than 50%, some even less than 25% (Fig 3B). Notably, only some effectors from Seq.C_72 have sequence annotation from SUPERFAMILY, identifying them as a Tubby C-terminal domain-like. The Tubby-like domain is characterized by a β -barrel structure enclosing an internalized α -helix in the center of the barrel. The predicted structure of Seq.C_72's effector in Struc.C_30, such as DK0911_02754, matches this structural feature. The remaining 6 sequence clusters within Struc.C_30 also exhibit high pTM scores. Furthermore, the representative proteins of these 6 sequence clusters show high structural similarity to the representative protein of Seq.C_72 (DK0911_02754) based on US-align analysis [79], indicating that they are all Tubby-like domain proteins despite some having low sequence similarity (22%-25%) with Seq.C_72 (Fig 3B) illustrating different effector sequences may adopt similar structures.

Further investigation revealed that Seq.C_133, Seq.C_191, and Seq.C_541, which are part of Struc.C_30, are also included in Struc.C_103. Proteins in Struc.C_103 are notably missing the complete central α -helix found in Struc.C_30, and some also lack parts of β -sheet in the β -barrel (Fig 3B). Despite including proteins from different sequence clusters, Struc.C_103 exhibits similar structural characteristics. This observation highlights that different sequences may adopt similar structural conformations. However, even though proteins from Seq.C_133, Seq.C_191, and Seq.C_541 share over 80% sequence similarity, their structural TM scores are relatively low and they are distributed into different structure clusters. The multiple sequence alignment resulted in the loss of central α -helix in Structure C_103, which is present in the Tubby-like effector candidates in Struc.C_30 (Figs 3 and S1). The sequence alignment clarifies the reason for the loss of α -helix. It shows deletion in the C-terminus of the effector candidates clustering in Struc.C_103. Although one effector from Struc.C_103, POW07379.1, kept a bit longer C-terminus sequence, but it lost the key amino acid cysteine to form a disulfide bond (S1 Fig). This observation suggests that these proteins have undergone different folding paths, potentially leading to divergent functions or loss of functional structural features.

Identified *Pst* effectors represent a new *Pst* effector structural family

To date, over 50 *Pst* effectors have been identified experimentally. We performed a BLASTP [80] alignment of these identified effectors against the mature sequences of 8,102 effector candidates, using a threshold of query coverage and percent identity $\geq 50\%$. This sequence alignment identified 547 effector candidates as sequence homology hits (S2 and S4 Tables). Using Foldseek and US-align for pair-wise comparison, we compared the predicted structures of 8,102 effector candidates with the identified *Pst* effectors and *Pgt*-Avr effectors (S2 Table), as well as with PDB chains, where there was only one 3D structure available for *Pst*, Pst_13661 (PDB chain: 8hf9_A), the others were of *Pgt*, AvrSr27 (PDB chain: 8v1j_A), AvrSr35 (PDB chain: 7xx2_B, 7xc2_D, 7xds_A, 7xds_B, 7xe0_B, 7xvg_B) and AvrSr50 (QCMJC) (PDB chain: 7mqq_A). In our data, structures with TM-scores greater than 0.5 by both Foldseek and US-align were homologous to the experimentally identified *Pst* effectors and *Pgt*-Avr, resulting in 3,707 effectors (S2 and S4 Tables). The sequence and structural comparisons are summarized in bubble plots (Fig 4A and 4B). Overall, structural comparisons enabled the identification of more effectors than sequence-based comparisons.

Pst_13661 is currently the only *Pst* effector with a resolved structure [34]. Based on the sequence of Pst_13661, homologous sequences were found in the races 104E137A-, PST-87/7, PST-78, and 93-210 (Fig 4A). In addition to the four aforementioned races, structurally similar homologs were also identified in *Pst* race PST-130 and *Ps* isolate 11-281 (Fig 4B). The predicted structures of the homologs in *Pst* race PST-130 and *Ps* isolate 11-281 are highly accurate and show significant similarity to the structure of Pst_13661 (PDB chain: 8hf9_A) (Fig 4B and 4C).

Interestingly, based on sequence similarity, PstGSRE4 homologs were found only in the effectors of *Pst* races PST-87/7 and PST-130, which belong to Struc.C_22 (Fig 4A). However, based on structural similarity, PstGSRE4, along with two other effectors, Pst27791 and PstSIE1, which do not share sequence similarity, showed predicted structural similarity to approximately 71% of the effectors in races or isolates belonging to Struc.C_22 and Struc.C_62 (Fig 4B). We then conducted a structural phylogenetic analysis of 125 effectors from 12 sequence clusters within Struc.C_22 and Struc.C_62, as well as the three identified effectors PstGSRE4, Pst27791, and PstSIE1 using DALI [81] (Fig 4D). This analysis assembled them into nine groups. Among these, the core structure of the second group, consisting of Pst104E_08415, PST877_18962, and PST43_09944, was formed by three helices, while the other groups had a core structure of four helices. Previous studies have reported that the host wheat interactors for Pst27791, PstGSRE4, and PstSIE1 are TaRaf46 [82], TaCZSOD2 [83], TaGAPDH2 [84], and TaSGT1 [85], respectively, which are key proteins in wheat immune pathways. This led us to identify a class of widely present effectors in *Pst* with a core structure of four helices that commonly interact with key immune pathway proteins during the infection of wheat.

The structure of *Pst* effector candidates from multiple different sequence clusters is similar to AvrSr35 and AvrSr50

Previous research predicted several *Pst* Avr candidates, including 48 secreted proteins and 14 non-secreted proteins [86]. We compared the sequences and the predicted structures of these 62 *Pst* Avr candidates with the sequences and predicted structures of 8,102 effector candidates (Fig 5A, S4 and S6 Tables). None of the non-secreted *Pst* Avr candidates showed sequence or predicted structural similarity to any of the effector candidates supporting our prediction that the effectors originate from the secretome. To date, no *Pst* Avr has been cloned and identified. However, in the closely related wheat rust pathogen *Puccinia graminis* f. sp. *tritici* (*Pgt*), five AvrS have been cloned and identified, they are AvrSr50, AvrSr35, AvrSr22, AvrSr13, AvrSr27. None of these *Pgt* AvrS showed sequence similarity to the 8,102 effectors candidates, AvrSr22, AvrSr13, and AvrSr27 nor did they show structural similarity. Notably, AvrSr35 and AvrSr50 exhibited structural similarity to several effectors with different sequences (Fig 5B and 5C and S2 and S4 Tables). Among them, effector candidates distributed in Struc.C_2, Struc.C_12, Struc.C_17, Struc.C_18, and Struc.C_26 from 26 sequence clusters are structurally AvrSr35-like, with a concentration in Struc.C_2. Effector candidates distributed in Struc.C_21 from 6 sequence clusters are structurally AvrSr50-like.

Pst effector candidates are structurally similar to some effectors from bacteria, oomycetes, and other fungi

In addition to *Pst* and *Pgt*, the structures of approximately 180 effectors from various bacteria, oomycetes, and fungi have been determined (S1 Table). None of these effectors showed sequence similarity to the 8,102 effector candidates using BlastP, indicating the distant phylogenetic relationships between these pathogens and *Pst*. However, some of these effectors exhibited structural similarities to the *Pst* effector candidates (Fig 6A and S1 and S4 Tables).

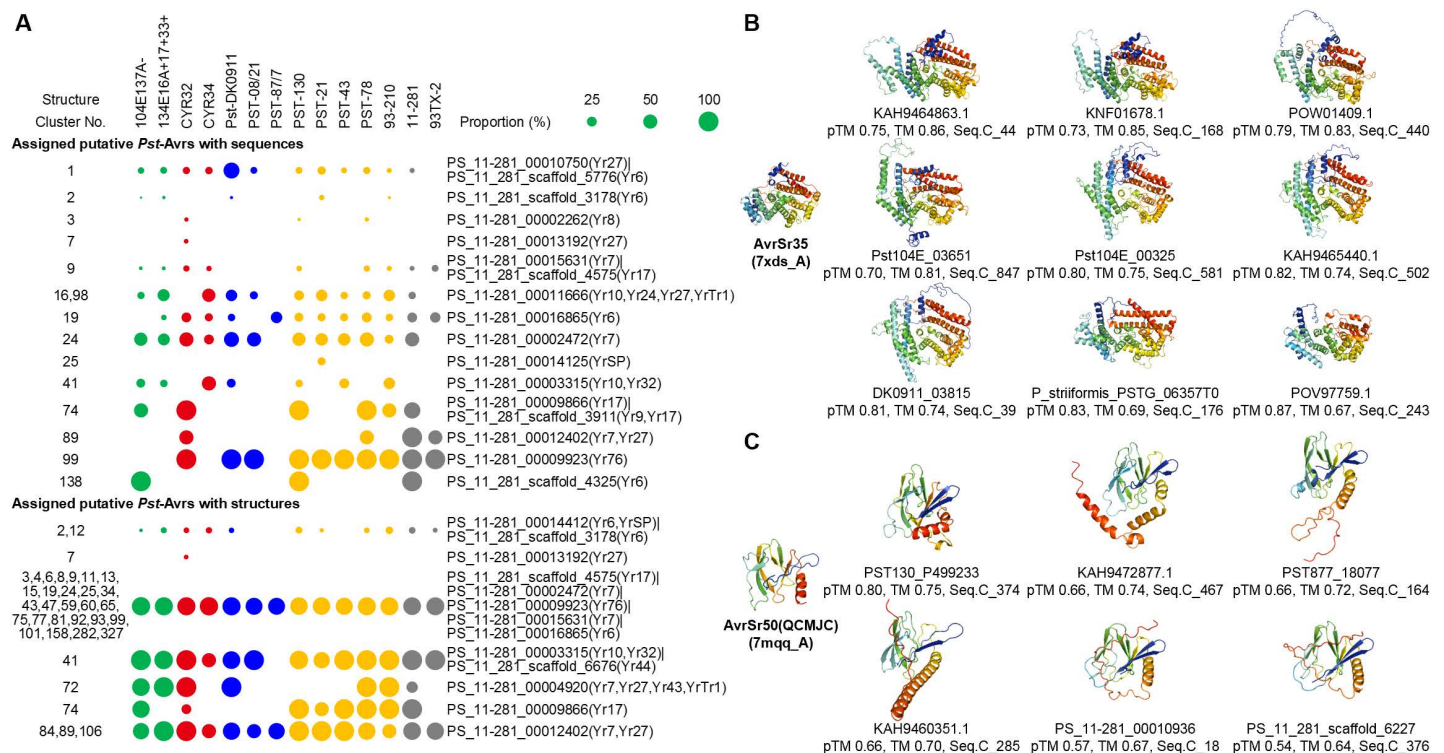


Fig 5. Sequence-based and structural analysis of putative *Pst* Avr candidates. (A) The proportion of putative *Pst* Avr candidates assigned to effector candidates from 14 *Pst* races or isolates within structure clusters, based on sequence and predicted structure comparison, shown with circles of varying sizes. (B, C) Structural analysis of AvrSr35 (7XDS_A) and AvrSr50(QCMJC) (7MQQ_A), along with predicted structures of AvrSr35-like and AvrSr50-like *Pst* effectors from different sequence clusters. The pTM values from AF2 modeling, TM-scores (TM) compared with AvrSr35 (7XDS_A) or AvrSr50(QCMJC) (7MQQ_A) structures, and sequence cluster numbers (Seq.C) are indicated below the protein IDs. Structures are shown from N-terminus (blue) to C-terminus (red).

<https://doi.org/10.1371/journal.pcbi.1012503.g005>

Several effectors characterized by two helices, such as AvrPto, AvrRps4C, Dld1, AVRvnt1, PHYLL_{OP}, and PHYLL_{PnWB}, showed structural similarity to 14% of the effectors distributed across various structure clusters from all races or isolates. Additionally, 15% of the effectors from all races or isolates distributed in Struc.C_7, Struc.C_21, Struc.C_110, and Struc.C_357 exhibited structural similarities to effectors from the Tox-like structure effector family, such as Avr2 (SIX3), PtrToxA, and SIX8. Some effectors from Struc.C_91 and Struc.C_271 also showed structural similarity to effectors from the MAX structure effector family, including PtrToxB, MAX28, and MoToxB.

Interestingly, Zt-KP4-1 showed structural similarity to 48% of the effectors distributed in Struc.C_1 and Struc.C_10 from all races or isolates. Among the effectors in Struc.C_1 and Struc.C_10 with structural similarity to Zt-KP4-1, some also exhibited sequence or structural similarity to identified *Pst* effectors Shr7 (PSTG_146695), Pst18363, PSTG_10917, and Pst_9302. Zt-KP4-1 indeed shared structural similarity with these identified *Pst* effectors (Fig 6B). Notably, Struc.C_1 contains the highest number of effectors in this study, suggesting that Zt-KP4-1-like structures are widespread in *Pst*. Additionally, we observed effectors such as PST130_P498983 with a tandem Zt-KP4-1-like structure and POW04647.1 with a triplet Zt-KP4-1-like structure (Fig 6B), indicating an expansion in the evolution of *Pst* effectors. Most significantly, MoHrip2 showed high structural similarity to all effectors in Struc.C_66 from all races or isolates and was exclusively similar to effectors in Struc.C_66 (Fig 6A and 6C).

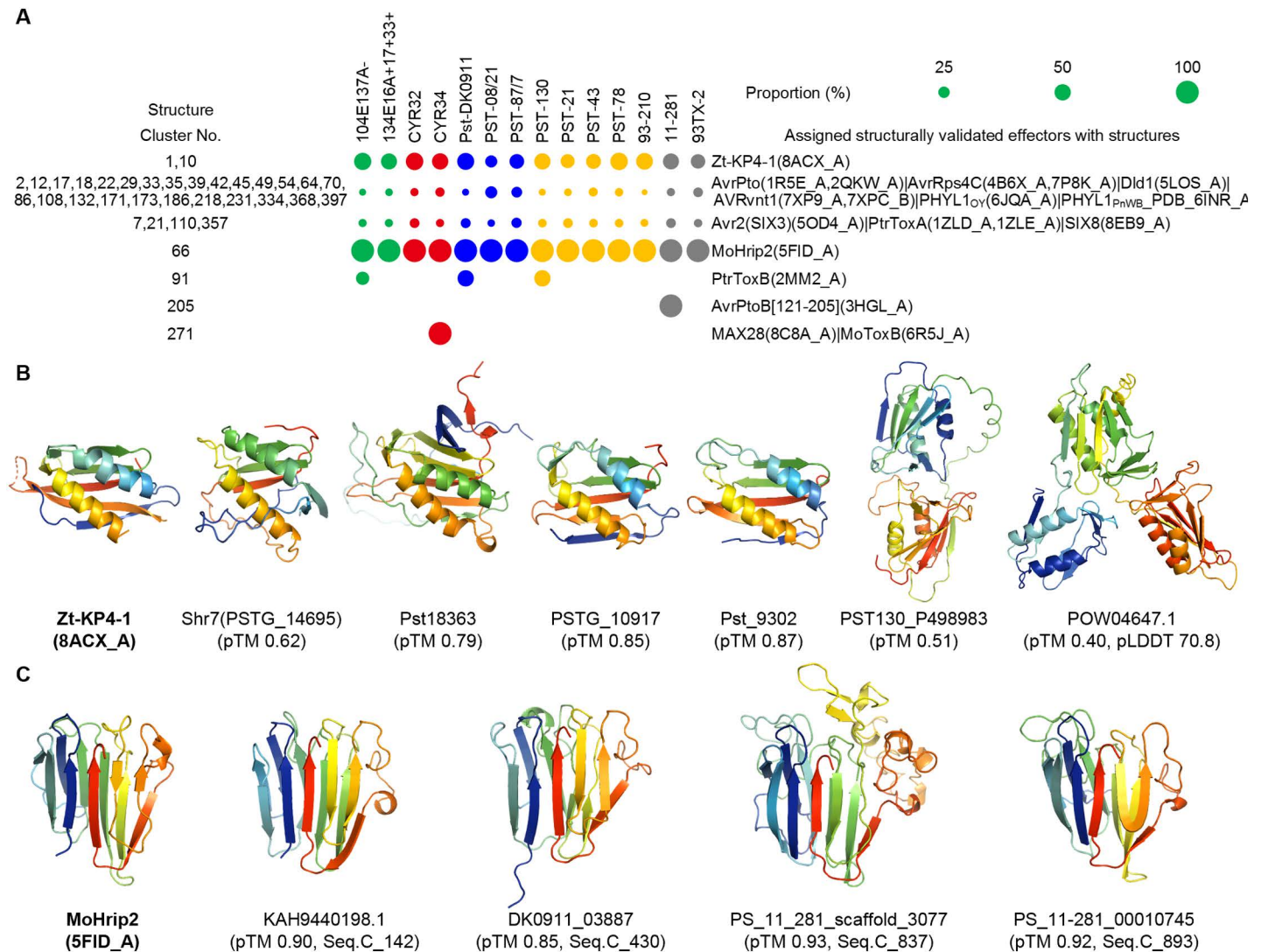


Fig 6. Structural analysis between structurally validated effectors and *Pst* effector candidates. (A) The proportion of structurally validated effectors assigned to the predicted structure of effector candidates from 14 *Ps* races or isolates within structure clusters, shown with circles of varying sizes. (B, C) Structural analysis of Zt-KP4-1 (8ACX_A) and MoHrip2 (5FID_A), along with representative predicted structures of Zt-KP4-1-like and MoHrip2-like *Pst* effectors. The pTM values from AF2 modeling are indicated below the protein IDs. For MoHrip2-like effectors, the sequence cluster number (Seq.C) is also indicated. Structures are shown from N-terminus (blue) to C-terminus (red).

<https://doi.org/10.1371/journal.pcbi.1012503.g006>

Discussion

Our comprehensive and detailed analysis of *Pst* effector candidates sheds light on the landscape of characterizations and the evolutionary dynamics of *Pst* effector repertoire, this study provides a detailed description of the diverse folds adopted by *Pst* effector proteins, offering new insights into their structural diversity. The identification of these folds may inform future studies on the potential molecular roles of *Pst* proteins and contribute to a better understanding of their involvement in stripe rust pathogenesis. We have collected a large number of *Pst* effector candidates for sequence and structural annotation, providing an essential resource for subsequent *Pst* effector research. This study is the first to perform structural clustering of *Pst* effector candidates examining their relationships from a structural perspective. We have

also explored and discovered for the first time, from a structural viewpoint, the existence of effector structural families among *Pst* effector candidates and their structural similarities with known effectors from other pathogens. This research offers valuable insights for the study of effectors of different pathogens, demonstrating how large-scale sequence and structural analyses can elucidate effector characteristics and advance effector research. Previous reviews or articles have summarized effectors with known structures [2,23]. Building on this foundation, we have further enhanced and supplemented the data, ultimately identifying over 180 structures (S1 Table). These findings will serve as a crucial training dataset for future research on effector structure, structural prediction, and structural family classification.

AlphaFold2 facilitates effector research

AlphaFold2 has revolutionized the field of protein structure prediction by providing highly accurate models for a vast array of proteins. However, when faced with proteins that lack homologous structures or have low sequence similarity to known templates, the accuracy of AlphaFold2's predictions can significantly diminish, leading to less reliable models for these novel proteins. Although this study focused on the sequence and structural analysis of 8,102 *Pst* effector candidates with well-predicted structures, it is important to note that nearly half of the predicted effectors were not included in this research due to poor structural predictions (pTM scores < 0.5 and pLDDT < 70). One possible reason for this is the strong sequence specificity of these *Pst* effectors, resulting in fewer reference templates during AF2 model construction, which hampers accurate modeling. However, this suggests that these effectors may have a higher specificity to *Pst*, potentially making them key effectors in the successful infection of wheat. For effectors with reference models, such as Pst_13661-like effectors, reliable models with pTM scores above 0.79 were obtained, demonstrating the potential for high-confidence predictions. Despite the capabilities of AF2 as a protein structure prediction tool, structural experiments remain essential for analyzing highly specific effectors, expanding the number of template models, and increasing prediction accuracy. AF2, while being a leading AI protein structure prediction tool, also has limitations. It cannot account for the cellular content where protein functions, such as pH, salt concentration, ions, and post-translational modifications, which are critical for protein conformation [87]. The newer AlphaFold3 has addressed some of these issues, such as introducing ions and ligands in structural predictions. Although there is substantial room for improvement in AF2, large-scale research on effector proteins based on their structures has significantly advanced. For example, understanding the structure of effectors can aid in the design of compounds as effector inhibitors [34]. AF2 can also facilitate protein interaction studies [88,89], greatly aiding in the exploration of pathogen-host interactions through the study of effector-interactor mechanisms [90].

Structural annotations assist effector characterizations

Before conducting effector cloning and functional identification, researchers often use protein sequence functional annotation databases to preliminarily predict the biological functions of effectors and obtain corresponding research direction ideas. These predictions are often unsatisfactory. It is important to note that the structural data used in this study are derived from computational predictions using AlphaFold2. While these predicted structures provide valuable insights and serve as a basis for clustering and annotation, they are not experimental ground truth. Variations in structural predictions, such as differences in disordered regions or loop conformations, may influence the clustering results and do not necessarily reflect true structural divergence. Experimental validation is essential to confirm these observations and refine our understanding of effector structure and function. In this study of 8,102 *Pst* effector

candidates, even after searching through 17 protein sequence functional or domain annotation databases, only about 21% of the effectors had functional or domain annotations based on their protein sequences. However, when we compared the predicted structures of *Pst* effector candidates with the PDB, CATH, and SCOP databases, and filtered out structural annotations with TM-scores < 0.5 , approximately 75% of the effector candidates still had structural annotation information. In this way, structure-guided similarity searches have made it possible to better annotate effector repertoires. Similar approaches have been successfully applied in works such as the AlphaFold clusters [91], UniProt3D [92], and the TED database [93], which focus on general protein annotations. Our study extends these methodologies specifically to the effector repertoire of *Pst*. Since structure often determines function, structural information can further predict or infer interactors within the host. The databases comparable in Foldseek are constantly being updated allowing comparisons not only with PDB, CATH, and SCOP but also with other databases to obtain more annotation information. Therefore, predicting the structures of *Pst* effectors and subsequently comparing them with protein structure annotation databases to obtain annotation information will provide preliminary ideas for the early stages of effector identification research. Our findings provide a detailed understanding of the structural and functional complexity of *Pst* effectors, offering a foundation for future studies on effector biology and host-pathogen interactions. These structural matches highlight similarities in overall fold families, but further analysis is required to establish functional relationships. Experimental validation and functional assays of these predicted effectors will be essential to fully understand their roles in *Pst* pathogenicity and host resistance.

Diverse sequences of effectors share a structural commonality

From the analysis of the relationship between the sequence and structure of effectors (Fig 1E and 1G and S5 Table), generally, effectors with similar sequences may adopt similar structures, aligning with our expectations. However, there are still instances where effectors with similar sequences form structures with low similarity. For example, in (Fig 3B), effectors belonging to the same sequence cluster with at least 82% sequence similarity form different structures and thus belong to different structure clusters. These observations are based on predicted structures, which may reflect conformational variability or artifacts rather than true structural divergence. Moreover, in addition to using sequences to identify homologs, structural predictions can also aid in determining homologs for *Pst* effectors, which has been tested in other organisms [94–96]. This study also found that effectors with different sequences can form similar structures. There are 129 structure clusters, which are ‘sequence-unrelated structurally similar’ (SUSS) clusters among 410 structure clusters, found in this study. This may arise from the pathogen using amino acid resources optimally during the infection process to form functionally similar effectors with similar structures.

New *Pst* effector families found

Although *Pst* cannot be cultured, and there is a lack of an efficient and reliable system for stable transformation, making it challenging to study its pathogenicity mechanisms through genetic methods, over 50 *Pst* effectors have already been identified (S2 Table). Previous research often studied effectors as isolated entities with little integration among them. So far, it has only been discovered that the identified *Pst* effectors Pst_12806, Pst_4, and Pst_5 share the same host wheat interactor TaISP [97,98] and that there is an interaction between PstCEP1 and PSTG_11208 [99]. In this study, by comparing the sequences and predicted structures of identified *Pst* effectors with 8,102 *Pst* effector candidates, we found that many identified *Pst* effectors have a widespread sequence or structural homologs in different races

or isolates. Furthermore, we discovered a more typical class of *Pst* effectors with a core structure of four helices, represented by the identified *Pst* effectors Pst27791, PstGSRE4, and PstSIE1, forming a unique *Pst* effector structural family. It has been demonstrated that *Pst* effectors Pst27791, PstGSRE4, and PstSIE1 play crucial roles in suppressing wheat's defense mechanisms during *Pst*-wheat interaction [82–85]. They are secreted and translocated into the cytoplasm of host cells, where they target their interactors. Pst27791 targets the Raf-like kinase TaRaf46 to inhibit ROS accumulation, MAPK activation, and defense-related gene expression [82]. PstGSRE4 suppresses the host defense response by targeting TaCZSOD2, inhibiting its enzyme activity to disrupt ROS-mediated hypersensitive response (HR) and disease resistance [83]. Additionally, PstGSRE4 also targets and stabilizes TaGAPDH2, further hindering the wheat defense system [84]. PstSIE1 targets TaSGT1 in wheat cells, interfering with the TaRAR1–TaSGT1 subcomplex formation to suppress defense responses [85]. In this study, structural predictions of these three experimentally identified *Pst* effectors and more *Pst* effector candidates revealed a conserved core structure of four α -helices with diverse functions. This finding provides a new perspective for future effector research, suggesting that by studying effector structural families comprehensively, we can better understand how these structural families of effectors function in host infection. It may be interesting to test the silencing of the effectors having highly similar structures, in addition to silencing the alleles to determine the function of any given effector. We have also observed that Struc.C_7, which contains 206 effector candidates, lacks annotation information and does not exhibit sequence or structural similarity to any of the identified effectors. However, as a structure cluster of effectors widely present in *Pst*, it remains uncharacterized and requires further study.

Potential *Pst* Avr candidates and likelihood of convergent evolutionary strategies

Identifying *Pst* Avr genes is crucial for understanding *Pst* variability. Although there have been reports predicting *Pst* Avr candidates, no *Pst* Avr has been identified to date. In this study, by comparing the predicted structures of *Pst* effector candidates with those of *Pgt* Avr, we identified many effectors that are AvrSr35-like and AvrSr50-like. These AvrSr35-like and AvrSr50-like *Pst* effector candidates could potentially be cognate Avr candidates for particular yellow rust resistance proteins and can be further experimentally validated. We also discovered that pathogens employ convergent evolutionary strategies for their effectors. Specifically, we found that the predicted structures of *Pst* effector candidates show high structural similarity to effectors from bacteria, oomycetes, and other fungi, despite having no sequence similarity in BlastP comparisons. In this study, at least 5.3% of the effector candidates are Zt-KP4-1-like and are primarily distributed in Struc.C_1 and Struc.C_10, indicating a commonality among *Pst* effectors.

The widespread presence of Zt-KP4-1-like structures among *Pst* effectors points to potential conserved mechanisms in effector evolution and function. Furthermore, the identification of structural similarities with known effectors from other pathogens, despite low sequence similarity, highlights the importance of structural analysis in uncovering functional relationships. These similarities suggest that effectors from different pathogens might converge on similar host targets or pathways, offering potential cross-species insights into effector biology.

Materials and Methods

Collection of *Puccinia striiformis* proteome and identified effectors and putative *Pst* Avr candidates

Proteomes of 14 races or isolates for *Puccinia striiformis* consisting of 357,396 proteins were collected (S3 Table). 58 identified *Pst* effectors and 5 avirulence factors (Avr) from stem rust

were collected from the literature and our lab, respectively (S2 Table). The 181 known effector structures were downloaded from PDB (S1 Table) (<https://www.rcsb.org/>, downloaded on 2024-06-13); only one (Pst_13661) is from *Puccinia striiformis* f.sp. *tritici*, all the other ones are from other plant pathogens. The putative *Pst* Avr of 62, i.e., 48 secreted and 14 non-secreted (S6 Table) were collected from Li et al., 2020 [86].

Effector prediction

SignalP 6.0 was used to identify the secreted proteins [46], glycosyl-phosphatidyl-inositol (GPI) anchoring containing proteins excluded with the help of NetGPI 1.1 [48]. The remaining secretome candidates were excluded if transmembrane was found with InterProScan 5.63-95.0 or if their signal peptides overlapped with PFAM domains over ten or more amino acids [47]. To determine the effectors, EffectorP 3.0 was used for effector prediction, including their cytoplasmic or apoplastic localization in the host [49].

Motifs analyses and subcellular localization prediction

Common effector motifs from oomycetes and fungi were searched, including RxLR [100], and YxSL[R/K] detected in oomycetes [101], [L/I]xAR and [R/K]CxxCx12H in some effectors of *Magnaporthe oryzae* [102], [R/K]VY[L/I]R identified in *Blumeria graminis* f. sp. *hordei* [103], [Y/F/W]xC found in the wheat powdery mildew [104] and rust effector candidates and G[I/F/Y] [A/L/S/T]R in some effectors of *Melampsora lini* [105]. ApoplastP [56], LOCALIZER [57], WoLF PSORT [58] and TargetP 2.0 [59], were used to predict the subcellular localization of effectors.

Effector characterization

Full-length sequences of effectors were used to search in the InterProScan 5.63–95.0 [60] against all databases available, i.e., PANTHER [60], SUPERFAMILY [62], Gene3D [63], Coils [64], ProSite Patterns and ProSite Profiles [65], CDD [66], FunFam [69], SMART [70], PRINTS [71], NCBIfam [73], Pfam [74], Hamap [75], PIRSF [76]. To find the cysteine residue count, mature sequences of effectors were used. MEROPS [67] was used to find the peptidases; for this purpose, HMMER was employed with the MEROPS database against our sequences (<https://www.ebi.ac.uk/Tools/hmmer/search/phmmer>, accessed on 2024-4-25). CAZy (Carbohydrate-Active Enzymes Database) term annotation performed with mature effector sequences in eggNOG-mapper 2.1 (accessed on 2024-4-24) [72]. KEGG Orthology Search was conducted on KofamKOALA (accessed on 2024-4-24) [68].

Structure prediction

The structures of 17,158 putative *Pst* effectors, using their mature sequences, were predicted by AlphaFold2 via the LocalColabFold approach [50,51]. Additionally, the structures of identified effectors, avirulence factors of stem rust, and putative *Pst* Avr candidates (S1, S2, and S6 Tables) were also predicted using the same method to efficiently utilize our resources. The computational workload for LocalColabFold was efficiently managed using the Lenovo ThinkBook 16p Gen 4, equipped with an Nvidia RTX4090 graphics card and the 13th Gen Intel Core i9 processor. Five models were generated, and the ranked_1 model was chosen as it had the best pLDDT score. All structures were then filtered based on pLDDT and pTM scores, retaining those with a pLDDT of 70 or above and/or a pTM of 0.5 or above for further analysis.

Clusters analyses

To create sequence clusters, CD-HIT was used with the sequence identity threshold set to 0.5 [52]. For structure clustering, the locally installed Foldseek (release 8-ef4e960) easy-cluster

option was used with a 0.5 threshold of sequence alignment coverage to group similar structures into the same cluster [53].

Structural annotation

We used Foldseek [53] for structural similarity search against the PDB chain, SCOPe40 and CATH50 databases [77,78,106], downloaded from Foldseek on 22 March 2024. Homologs with a TM score greater than 0.5 were retained, with a maximum of ten hits per query.

Network analysis

We selected 1,178 representative structures with the longest sequence length of each sequence cluster within each structure cluster from our dataset (marked in grey in S4 Table) and computed their pair-wise TM scores using Foldseek. Those edges with TM scores > 0.5 were then imported into Gephi 0.10.1 with the layout of Fruchterman-Reingold to construct and visualize our network for further analysis.

Homologous effectors search

We used Foldseek and US-align [79] tools to compare predicted protein structures from various sources, i.e., 58 identified *Pst* effectors, 5 Avr from stem rust, 62 putative *Pst*-Avrs, and 181 effector structures downloaded from the PDB. We identified matches with a TM score of 0.5 or higher as potentially homologous. We utilized BlastP [80] to identify sequence similarities among the proteins mentioned above and considered matches where the query coverage and percent identity were both 50% or higher.

Program and software

We used TBtools-II [107] for converting FASTA files to tables and running InterProScan analyses in a loop. Origin 2022 facilitated the creation of violin plots, histograms, and stacked bar charts. DALI was utilized to generate Newick dendrograms for structural comparisons [81]. iTOL (<https://itol.embl.de/>) was used for visualizing phylogenetic tree analyses. Cluster heatmap and PCA (principal component analysis) were performed using the OECloud tools at <https://cloud.oebiotech.com>. Protein structures were visualized and edited in the PyMOL [108]. Multiple sequence alignment was performed with ESPript 3 (<https://espript.ibcp.fr/ESPrript/ESPrript/>) [109]. Pairwise sequence similarity analysis in each structural cluster was analysed by CLUSTALW (<https://www.genome.jp/tools-bin/clustalw>).

Supporting information

S1 Fig. The multiple sequence alignment of Tubby-like effector candidates showing in the Fig 3B. The secondary structure features showing above the alignments from the AlphaFold 2 predicted structure of DK0911_02754. The C-terminus of the sequences with blue background indicates the Tubby-like effector candidates from structure cluster No. 30 (Struc.C_30). The C-terminus of the sequences with green background indicates the Tubby-like effector candidates from Struc.C_103. The corresponding position of 68 Cysteine and 189 Cysteine in the sequence of DK0911_02754 indicates the formation of disulfide bond, marking in green '1' below the alignments.
(DOCX)

S1 Table. Identified and structurally resolved effectors or avirulence factors of plant pathogens with their structure homologous analyses.
(XLSX)

S2 Table. Identified *Puccinia striiformis* f. sp. *tritici* effectors and avirulence factors of *Puccinia graminis* f. sp. *tritici* with their homologous analyses.

(XLSX)

S3 Table. Protein sets of 14 *Puccinia striiformis* races or isolates used for analyses and statistics from proteome to effectors.

(XLSX)

S4 Table. The metadata for 8,102 predicted effectors with well-fold from 14 *Puccinia striiformis* races or isolates.

(XLSX)

S5 Table. Statistics between sequence cluster and structure cluster of *Puccinia striiformis* f. sp. *tritici* effector candidates.

(XLSX)

S6 Table. Putative avirulence factors of *Puccinia striiformis* f. sp. *tritici* with their homologous analyses.

(XLSX)

Acknowledgements

We would like to thank Chaoming Zhang for his assistance in predicting the protein structures for specific races using AlphaFold2. We also extend our gratitude to Diane G. O. Saunders and Cristobal Uauy for providing the proteome data of PST-21, PST-43, PST-87/7, and PST-08/21. Additionally, we sincerely thank Dr. Shozeb Haider for his valuable support. Their contributions were invaluable to our research.

Author contributions

Conceptualization: Mahinur Akkaya.

Data curation: Raheel Asghar, Nan Wu, Yulei Wang.

Formal analysis: Raheel Asghar, Nan Wu.

Investigation: Raheel Asghar, Nan Wu.

Methodology: Raheel Asghar, Nan Wu, Mahinur Akkaya.

Resources: Mahinur Akkaya.

Software: Raheel Asghar, Nan Wu, Noman Ali.

Supervision: Mahinur Akkaya.

Validation: Nan Wu, Mahinur Akkaya.

Writing – original draft: Raheel Asghar, Nan Wu, Mahinur Akkaya.

Writing – review & editing: Raheel Asghar, Nan Wu, Noman Ali, Yulei Wang, Mahinur Akkaya.

References

1. Toruño TY, Stergiopoulos I, Coaker G. Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. *Annu Rev Phytopathol.* 2016;54:419–41. <https://doi.org/10.1146/annurev-phyto-080615-100204> PMID: 27359369
2. Mukhi N, Gorenkin D, Banfield MJ. Exploring folds, evolution and host interactions: understanding effector structure/function in disease and immunity. *New Phytol.* 2020;227(2):326–33. <https://doi.org/10.1111/nph.16563> PMID: 32239533

3. 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;81(2):e00066-16. <https://doi.org/10.1128/MMBR.00066-16> PMID: [28356329](#)
4. Outram MA, Figueroa M, Sperschneider J, Williams SJ, Dodds PN. Seeing is believing: Exploiting advances in structural biology to understand and engineer plant immunity. *Curr Opin Plant Biol*. 2022;67:102210. <https://doi.org/10.1016/j.pbi.2022.102210> PMID: [35461025](#)
5. Lovelace AH, Dorhmi S, Hulin MT, Li Y, Mansfield JW, Ma W. Effector Identification in Plant Pathogens. *Phytopathology*. 2023;113(4):637–50. <https://doi.org/10.1094/PHYTO-09-22-0337-KD> PMID: [37126080](#)
6. Boutemy LS, King SRF, Win J, Hughes RK, Clarke TA, Blumenschein TMA, et al. Structures of Phytophthora RXLR effector proteins: a conserved but adaptable fold underpins functional diversity. *J Biol Chem*. 2011;286(41):35834–42. <https://doi.org/10.1074/jbc.M111.262303> PMID: [21813644](#)
7. He J, Ye W, Choi DS, Wu B, Zhai Y, Guo B, et al. Structural analysis of Phytophthora suppressor of RNA silencing 2 (PSR2) reveals a conserved modular fold contributing to virulence. *Proc Natl Acad Sci U S A*. 2019;116(16):8054–9. <https://doi.org/10.1073/pnas.1819481116> PMID: [30926664](#)
8. Guncar G, Wang C-IA, Forwood JK, Teh T, Catanzariti A-M, Ellis JG, et al. The use of Co2+ for crystallization and structure determination, using a conventional monochromatic X-ray source, of flax rust avirulence protein. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2007;63(Pt 3):209–13. <https://doi.org/10.1107/S1744309107004599> PMID: [17329816](#)
9. Wang C-IA, Guncar G, Forwood JK, Teh T, Catanzariti A-M, Lawrence GJ, et al. Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *Plant Cell*. 2007;19(9):2898–912. <https://doi.org/10.1105/tpc.107.053611> PMID: [17873095](#)
10. Di X, Cao L, Hughes RK, Tintor N, Banfield MJ, Takken FLW. Structure-function analysis of the Fusarium oxysporum Avr2 effector allows uncoupling of its immune-suppressing activity from recognition. *New Phytol*. 2017;216(3):897–914. <https://doi.org/10.1111/nph.14733> PMID: [28857169](#)
11. Yu DS, Outram MA, Smith A, McCombe CL, Khambalkar PB, Rima SA, et al. The structural repertoire of Fusarium oxysporum f. sp. lycopersici effectors revealed by experimental and computational studies. *Elife*. 2024;12:RP89280. <https://doi.org/10.7554/eLife.89280> PMID: [38411527](#)
12. Sarma GN, Manning VA, Ciuffetti LM, Karplus PA. Structure of Ptr ToxA: an RGD-containing host-selective toxin from Pyrenophora tritici-repentis. *Plant Cell*. 2005;17(11):3190–202. <https://doi.org/10.1105/tpc.105.034918> PMID: [16214901](#)
13. 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;55(2):219–23. <https://doi.org/10.1007/s10858-012-9695-5> PMID: [23334361](#)
14. 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. *PLoS Pathog*. 2015;11(10):e1005228. <https://doi.org/10.1371/journal.ppat.1005228> PMID: [26506000](#)
15. Ose T, Oikawa A, Nakamura Y, Maenaka K, Higuchi Y, Satoh Y, et al. Solution structure of an avirulence protein, AVR-Pia, from Magnaporthe oryzae. *J Biomol NMR*. 2015;63(2):229–35. <https://doi.org/10.1007/s10858-015-9979-7> PMID: [26362280](#)
16. Nyarko A, Singarapu KK, Figueroa M, Manning VA, Pandelova I, Wolpert TJ, et al. Solution NMR structures of Pyrenophora tritici-repentis ToxB and its inactive homolog reveal potential determinants of toxin activity. *J Biol Chem*. 2014;289(37):25946–56. <https://doi.org/10.1074/jbc.M114.569103> PMID: [25063993](#)
17. Pennington HG, Jones R, Kwon S, Bonciani G, Thieron H, Chandler T, et al. The fungal ribonuclease-like effector protein CSEP0064/BEC1054 represses plant immunity and interferes with degradation of host ribosomal RNA. *PLoS Pathog*. 2019;15(3):e1007620. <https://doi.org/10.1371/journal.ppat.1007620> PMID: [30856238](#)
18. Cao Y, Kümmel F, Logemann E, Gebauer JM, Lawson AW, Yu D, et al. Structural polymorphisms within a common powdery mildew effector scaffold as a driver of coevolution with cereal immune receptors. *Proc Natl Acad Sci U S A*. 2023;120(32):e2307604120. <https://doi.org/10.1073/pnas.2307604120> PMID: [37523523](#)
19. Blondeau K, Blaise F, Graille M, Kale SD, Linglin J, Ollivier B, et al. Crystal structure of the effector AvrLm4-7 of Leptosphaeria maculans reveals insights into its translocation into plant cells and recognition by resistance proteins. *Plant J*. 2015;83(4):610–24. <https://doi.org/10.1111/tpj.12913> PMID: [26082394](#)
20. Lazar N, Mesarich CH, Petit-Houedet Y, Talbi N, Li de la Sierra-Gallay I, Zélie E, et al. A new family of structurally conserved fungal effectors displays epistatic interactions with plant resistance proteins. *PLoS Pathog*. 2022;18(7):e1010664. <https://doi.org/10.1371/journal.ppat.1010664> PMID: [35793393](#)

21. Derbyshire MC, Raffaele S. Surface frustration re-patterning underlies the structural landscape and evolvability of fungal orphan candidate effectors. *Nat Commun.* 2023;14(1):5244. <https://doi.org/10.1038/s41467-023-40949-9> PMID: 37640704
22. Rozano L, Mukuka YM, Hane JK, Mancera RL. Ab Initio Modelling of the Structure of ToxA-like and MAX Fungal Effector Proteins. *Int J Mol Sci.* 2023;24(7):6262. <https://doi.org/10.3390/ijms24076262> PMID: 37047233
23. Seong K, Krasileva KV. Prediction of effector protein structures from fungal phytopathogens enables evolutionary analyses. *Nat Microbiol.* 2023;8(1):174–87. <https://doi.org/10.1038/s41564-022-01287-6> PMID: 36604508
24. Yan X, Tang B, Ryder LS, MacLean D, Were VM, Eseola AB, et al. The transcriptional landscape of plant infection by the rice blast fungus *Magnaporthe oryzae* reveals distinct families of temporally co-regulated and structurally conserved effectors. *Plant Cell.* 2023;35(5):1360–85. <https://doi.org/10.1093/plcell/koad036> PMID: 36808541
25. De la Concepcion JC, Langner T, Fujisaki K, Yan X, Were V, Lam AHC, et al. Zinc-finger (ZiF) fold secreted effectors form a functionally diverse family across lineages of the blast fungus *Magnaporthe oryzae*. *PLoS Pathog.* 2024;20(6):e1012277. <https://doi.org/10.1371/journal.ppat.1012277> PMID: 38885263
26. Schuster M, Schweizer G, Reißmann S, Happel P, Aßmann D, Rössel N, et al. Novel secreted effectors conserved among smut fungi contribute to the virulence of *Ustilago maydis*. *Mol Plant Microbe Interact.* 2024;37(3):250–63. <https://doi.org/10.1094/MPMI-09-23-0139-FI> PMID: 38416124
27. Zhang Z, Zhang X, Tian Y, Wang L, Cao J, Feng H, et al. Complete telomere-to-telomere genomes uncover virulence evolution conferred by chromosome fusion in oomycete plant pathogens. *Nat Commun.* 2024;15(1):4624. <https://doi.org/10.1038/s41467-024-49061-y> PMID: 38816389
28. Chen X. Pathogens which threaten food security: *Puccinia striiformis*, the wheat stripe rust pathogen. *Food Sec.* 2020;12(2):239–51. <https://doi.org/10.1007/s12571-020-01016-z>
29. Schwessinger B. Fundamental wheat stripe rust research in the 21st century. *New Phytol.* 2017;213(4):1625–31. <https://doi.org/10.1111/nph.14159> PMID: 27575735
30. Xia C, Qiu A, Wang M, Liu T, Chen W, Chen X. Current Status and Future Perspectives of Genomics Research in the Rust Fungi. *Int J Mol Sci.* 2022;23(17):9629. <https://doi.org/10.3390/ijms23179629> PMID: 36077025
31. Wu N, Ozketen AC, Cheng Y, Jiang W, Zhou X, Zhao X, et al. *Puccinia striiformis* f. sp. tritici effectors in wheat immune responses. *Front Plant Sci.* 2022;13:1012216. <https://doi.org/10.3389/fpls.2022.1012216> PMID: 36420019
32. Wang J, Chen T, Tang Y, Zhang S, Xu M, Liu M, et al. The Biological Roles of *Puccinia striiformis* f. sp. tritici Effectors during Infection of Wheat. *Biomolecules.* 2023;13(6):889. <https://doi.org/10.3390/biom13060889> PMID: 37371469
33. Xu Q, Wang J, Zhao J, Xu J, Sun S, Zhang H, et al. A polysaccharide deacetylase from *Puccinia striiformis* f. sp. tritici is an important pathogenicity gene that suppresses plant immunity. *Plant Biotechnol J.* 2020;18(8):1830–42. <https://doi.org/10.1111/pbi.13345> PMID: 31981296
34. Liu L, Xia Y, Li Y, Zhou Y, Su X, Yan X, et al. Inhibition of chitin deacetylases to attenuate plant fungal diseases. *Nat Commun.* 2023;14(1):3857. <https://doi.org/10.1038/s41467-023-39562-7> PMID: 37385996
35. Lubega J, Figueroa M, Dodds PN, Kanyuka K. Comparative Analysis of the Avirulence Effectors Produced by the Fungal Stem Rust Pathogen of Wheat. *Mol Plant Microbe Interact.* 2024;37(3):171–8. <https://doi.org/10.1094/MPMI-10-23-0169-FI> PMID: 38170736
36. Chen J, Upadhyaya NM, Ortiz D, Sperschneider J, Li F, Bouton C, et al. Loss of AvrSr50 by somatic exchange in stem rust leads to virulence for Sr50 resistance in wheat. *Science.* 2017;358(6370):1607–10. <https://doi.org/10.1126/science.aao4810> PMID: 29269475
37. Ortiz D, Chen J, Outram MA, Saur IML, Upadhyaya NM, Mago R, et al. The stem rust effector protein AvrSr50 escapes Sr50 recognition by a substitution in a single surface-exposed residue. *New Phytol.* 2022;234(2):592–606. <https://doi.org/10.1111/nph.18011> PMID: 35107838
38. Salcedo A, Rutter W, Wang S, Akhunova A, Bolus S, Chao S, et al. Variation in the AvrSr35 gene determines Sr35 resistance against wheat stem rust race Ug99. *Science.* 2017;358(6370):1604–6. <https://doi.org/10.1126/science.aao7294> PMID: 29269474
39. Förderer A, Li E, Lawson AW, Deng Y-N, Sun Y, Logemann E, et al. A wheat resistosome defines common principles of immune receptor channels. *Nature.* 2022;610(7932):532–9. <https://doi.org/10.1038/s41586-022-05231-w> PMID: 36163289
40. Zhao Y-B, Liu M-X, Chen T-T, Ma X, Li Z-K, Zheng Z, et al. Pathogen effector AvrSr35 triggers Sr35 resistosome assembly via a direct recognition mechanism. *Sci Adv.* 2022;8(36):eabq5108. <https://doi.org/10.1126/sciadv.abq5108> PMID: 36083908

41. Arndell T, Chen J, Sperschneider J, Upadhyaya NM, Blundell C, Niesner N, et al. Pooled effector library screening in protoplasts rapidly identifies novel Avr genes. *Nat Plants*. 2024;10(4):572–80. <https://doi.org/10.1038/s41477-024-01641-y> PMID: 38409291
42. Upadhyaya NM, Mago R, Panwar V, Hewitt T, Luo M, Chen J, et al. Genomics accelerated isolation of a new stem rust avirulence gene-wheat resistance gene pair. *Nat Plants*. 2021;7(9):1220–8. <https://doi.org/10.1038/s41477-021-00971-5> PMID: 34294906
43. Outram MA, Chen J, Broderick S, Li Z, Aditya S, Tasneem N, et al. AvrSr27 is a zinc-bound effector with a modular structure important for immune recognition. *New Phytol*. 2024;243(1):314–29. <https://doi.org/10.1111/nph.19801> PMID: 38730532
44. Seong K, Krasileva KV. Computational Structural Genomics Unravels Common Folds and Novel Families in the Secretome of Fungal Phytopathogen *Magnaporthe oryzae*. *Mol Plant Microbe Interact*. 2021;34(11):1267–80. <https://doi.org/10.1094/MPMI-03-21-0071-R> PMID: 34415195
45. Rozano L, Jones DAB, Hane JK, Mancera RL. Template-Based Modelling of the Structure of Fungal Effector Proteins. *Mol Biotechnol*. 2024;66(4):784–813. <https://doi.org/10.1007/s12033-023-00703-4> PMID: 36940017
46. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsigirgos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol*. 2022;40(7):1023–5. <https://doi.org/10.1038/s41587-021-01156-3> PMID: 34980915
47. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. *Nucleic Acids Res*. 2023;51(D1):D418–27. <https://doi.org/10.1093/nar/gkac993> PMID: 36350672
48. Gíslason MH, Nielsen H, Almagro Armenteros JJ, Johansen AR. Prediction of GPI-anchored proteins with pointer neural networks. *Current Research in Biotechnology*. 2021;3:6–13. <https://doi.org/10.1016/j.crbiot.2021.01.001>
49. Sperschneider J, Dodds PN. EffectorP 3.0: Prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol Plant Microbe Interact*. 2022; 35(2):146–56. <https://doi.org/10.1094/MPMI-08-21-0201-R> PMID: 34698534
50. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2> PMID: 34265844
51. Mirdita M, Schütze K, Moriawaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods*. 2022;19(6):679–82. <https://doi.org/10.1038/s41592-022-01488-1> PMID: 35637307
52. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658–9. <https://doi.org/10.1093/bioinformatics/btl158> PMID: 16731699
53. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol*. 2024;42(2):243–6. <https://doi.org/10.1038/s41587-023-01773-0> PMID: 37156916
54. Kamoun S. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annu Rev Phytopathol*. 2006;44:41–60. <https://doi.org/10.1146/annurev.phyto.44.070505.143436> PMID: 16448329
55. Saunders DGO, Win J, Cano LM, Szabo LJ, Kamoun S, Raffaele S. Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS One*. 2012;7(1):e29847. <https://doi.org/10.1371/journal.pone.0029847> PMID: 22238666
56. Sperschneider J, Dodds PN, Singh KB, Taylor JM. ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. *New Phytol*. 2018;217(4):1764–78. <https://doi.org/10.1111/nph.14946> PMID: 29243824
57. Sperschneider J, Catanzariti A-M, DeBoer K, Petre B, Gardiner DM, Singh KB, et al. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Sci Rep*. 2017;7:44598. <https://doi.org/10.1038/srep44598> PMID: 28300209
58. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. *Nucleic Acids Res*. 2007;35(Web Server issue):W585–7. <https://doi.org/10.1093/nar/gkm259> PMID: 17517783
59. Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, et al. Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance*. 2019;2(5):e201900429. <https://doi.org/10.26508/lsa.201900429> PMID: 31570514
60. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou L-P, Mi H. PANTHER: making genome-scale phylogenetics accessible to all. *Protein Sci*. 2022;31(1):8–22. <https://doi.org/10.1002/pro.4218> PMID: 34717010

61. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. *Nucleic Acids Res.* 2021;49(D1):D412–9. <https://doi.org/10.1093/nar/gkaa913> PMID: 33125078
62. Oates ME, Stahlhacke J, Vavoulis DV, Smithers B, Rackham OJL, Sardar AJ, et al. The SUPERFAMILY 1.75 database in 2014: a doubling of data. *Nucleic Acids Res.* 2015;43(Database issue):D227–33. <https://doi.org/10.1093/nar/gku1041> PMID: 25414345
63. Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, et al. CATH: increased structural coverage of functional space. *Nucleic Acids Res.* 2021;49(D1):D266–73. <https://doi.org/10.1093/nar/gkaa1079> PMID: 33237325
64. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science.* 1991;252(5009):1162–4. <https://doi.org/10.1126/science.252.5009.1162> PMID: 2031185
65. Sigrist CJA, de Castro E, Cerutti L, Cuče BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. *Nucleic Acids Res.* 2013;41(Database issue):D344–7. <https://doi.org/10.1093/nar/gks1067> PMID: 23161676
66. Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, et al. The conserved domain database in 2023. *Nucleic Acids Res.* 2023;51(D1):D384–8. <https://doi.org/10.1093/nar/gkac1096> PMID: 36477806
67. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 2018;46(D1):D624–32. <https://doi.org/10.1093/nar/gkx1134> PMID: 29145643
68. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics.* 2020;36(7):2251–2. <https://doi.org/10.1093/bioinformatics/btz859> PMID: 31742321
69. Das S, Lee D, Sillitoe I, Dawson NL, Lees JG, Orengo CA. Functional classification of CATH superfamilies: a domain-based approach for protein function annotation. *Bioinformatics.* 2015;31(21):3460–7. <https://doi.org/10.1093/bioinformatics/btv398> PMID: 26139634
70. Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* 2021;49(D1):D458–60. <https://doi.org/10.1093/nar/gkaa937> PMID: 33104802
71. Attwood TK, Coletta A, Muirhead G, Pavlopoulou A, Philippou PB, Popov I, et al. The PRINTS database: a fine-grained protein sequence annotation and analysis resource—its status in 2012. *Database (Oxford).* 2012;2012bas019. <https://doi.org/10.1093/database/bas019> PMID: 22508994
72. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol.* 2021;38(12):5825–9. <https://doi.org/10.1093/molbev/msab293> PMID: 34597405
73. Haft DH, Badretdin A, Coulouris G, DiCuccio M, Durkin AS, Jovenitti E, et al. RefSeq and the prokaryotic genome annotation pipeline in the age of metagenomes. *Nucleic Acids Res.* 2024;52(D1):D762–9. <https://doi.org/10.1093/nar/gkad988> PMID: 37962425
74. Pedruzzi I, Rivoire C, Auchincloss AH, Coudert E, Keller G, de Castro E, et al. HAMAP in 2015: updates to the protein family classification and annotation system. *Nucleic Acids Res.* 2015;43(Database issue):D1064–70. <https://doi.org/10.1093/nar/gku1002> PMID: 25348399
75. Nikolskaya AN, Arighi CN, Huang H, Barker WC, Wu CH. PIRSF family classification system for protein functional and evolutionary analysis. *Evol Bioinform Online.* 2007;2:197–209. PMID: 19455212
76. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chao H, Chen L, et al. RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res.* 2023;51(D1):D488–508. <https://doi.org/10.1093/nar/gkac1077> PMID: 36420884
77. Fox NK, Brenner SE, Chandonia J-M. SCOPe: structural classification of proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res.* 2014;42(Database issue):D304–9. <https://doi.org/10.1093/nar/gkt1240> PMID: 24304899
78. Chandonia J-M, Guan L, Lin S, Yu C, Fox NK, Brenner SE. SCOPe: improvements to the structural classification of proteins - extended database to facilitate variant interpretation and machine learning. *Nucleic Acids Res.* 2022;50(D1):D553–9. <https://doi.org/10.1093/nar/gkab1054> PMID: 34850923
79. Zhang C, Shine M, Pyle AM, Zhang Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat Methods.* 2022;19(9):1109–15. <https://doi.org/10.1038/s41592-022-01585-1> PMID: 36038728

80. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 2008;36(Web Server issue):W5-9. <https://doi.org/10.1093/nar/gkn201> PMID: [18440982](https://pubmed.ncbi.nlm.nih.gov/18440982/)
81. Holm L. Dali server: structural unification of protein families. *Nucleic Acids Res.* 2022;50(W1):W210–5. <https://doi.org/10.1093/nar/gkac387> PMID: [35610055](https://pubmed.ncbi.nlm.nih.gov/35610055/)
82. Wan C, Liu Y, Tian S, Guo J, Bai X, Zhu H, et al. A serine-rich effector from the stripe rust pathogen targets a Raf-like kinase to suppress host immunity. *Plant Physiol.* 2022;190(1):762–78. <https://doi.org/10.1093/plphys/kiac218> PMID: [35567492](https://pubmed.ncbi.nlm.nih.gov/35567492/)
83. Liu C, Wang Y, Wang Y, Du Y, Song C, Song P, et al. Glycine-serine-rich effector PstGSRE4 in *Puccinia striiformis* f. sp. tritici inhibits the activity of copper zinc superoxide dismutase to modulate immunity in wheat. *PLoS Pathog.* 2022;18(7):e1010702. <https://doi.org/10.1371/journal.ppat.1010702> PMID: [35881621](https://pubmed.ncbi.nlm.nih.gov/35881621/)
84. Liu C, Wang Y, Du Y, Kang Z, Guo J, Guo J. Glycine-serine-rich effector PstGSRE4 in *Puccinia striiformis* f. sp. tritici targets and stabilizes TaGAPDH2 that promotes stripe rust disease. *Plant Cell Environ.* 2024;47(3):947–60. <https://doi.org/10.1111/pce.14786> PMID: [38105492](https://pubmed.ncbi.nlm.nih.gov/38105492/)
85. Wang Y, Liu C, Du Y, Cai K, Wang Y, Guo J, et al. A stripe rust fungal effector PstSIE1 targets TaSGT1 to facilitate pathogen infection. *Plant J.* 2022;112(6):1413–28. <https://doi.org/10.1111/tpj.16019> PMID: [36308427](https://pubmed.ncbi.nlm.nih.gov/36308427/)
86. Li Y, Xia C, Wang M, Yin C, Chen X. Whole-genome sequencing of *Puccinia striiformis* f. sp. tritici mutant isolates identifies avirulence gene candidates. *BMC Genomics.* 2020;21(1):247. <https://doi.org/10.1186/s12864-020-6677-y> PMID: [32197579](https://pubmed.ncbi.nlm.nih.gov/32197579/)
87. Wang L, Wen Z, Liu S-W, Zhang L, Finley C, Lee H-J, et al. Overview of alphafold2 and breakthroughs in overcoming its limitations. *Comput Biol Med.* 2024;176:108620. <https://doi.org/10.1016/j.compbimed.2024.108620> PMID: [38761500](https://pubmed.ncbi.nlm.nih.gov/38761500/)
88. Homma F, Huang J, van der Hoorn RAL. AlphaFold-Multimer predicts cross-kingdom interactions at the plant-pathogen interface. *Nat Commun.* 2023;14(1):6040. <https://doi.org/10.1038/s41467-023-41721-9> PMID: [37758696](https://pubmed.ncbi.nlm.nih.gov/37758696/)
89. Wang L, Jia Y, Osakina A, Olsen KM, Huang Y, Jia MH, et al. Receptor- ligand interactions in plant innate immunity revealed by AlphaFold protein structure prediction. *bioRxiv.* 2024. <https://doi.org/10.1101/2024.06.12.598632>
90. Jones JDG, Staskawicz BJ, Dangl JL. The plant immune system: From discovery to deployment. *Cell.* 2024;187(9):2095–116. <https://doi.org/10.1016/j.cell.2024.03.045> PMID: [38670067](https://pubmed.ncbi.nlm.nih.gov/38670067/)
91. Barrio-Hernandez I, Yeo J, J  nes J, Mirdita M, Gilchrist CLM, Wein T, et al. Clustering predicted structures at the scale of the known protein universe. *Nature.* 2023;622(7983):637–45. <https://doi.org/10.1038/s41586-023-06510-w> PMID: [37704730](https://pubmed.ncbi.nlm.nih.gov/37704730/)
92. Durairaj J, Waterhouse AM, Mets T, Brodiazenko T, Abdullah M, Studer G, et al. Uncovering new families and folds in the natural protein universe. *Nature.* 2023;622(7983):646–53. <https://doi.org/10.1038/s41586-023-06622-3> PMID: [37704037](https://pubmed.ncbi.nlm.nih.gov/37704037/)
93. Lau AM, Bordin N, Kandathil SM, Sillitoe I, Waman VP, Wells J, et al. Exploring structural diversity across the protein universe with The Encyclopedia of Domains. *Science.* 2024;386(6721):eadq4946. <https://doi.org/10.1126/science.adq4946> PMID: [39480926](https://pubmed.ncbi.nlm.nih.gov/39480926/)
94. Gligorijevi   V, Renfrew PD, Kosci  lek T, Leman JK, Berenberg D, Vatanen T, et al. Structure-based protein function prediction using graph convolutional networks. *Nat Commun.* 2021;12(1):3168. <https://doi.org/10.1038/s41467-021-23303-9> PMID: [34039967](https://pubmed.ncbi.nlm.nih.gov/34039967/)
95. Hamamsy T, Morton JT, Blackwell R, Berenberg D, Carriero N, Gligorijevic V, et al. Protein remote homology detection and structural alignment using deep learning. *Nat Biotechnol.* 2024;42(6):975–85. <https://doi.org/10.1038/s41587-023-01917-2> PMID: [37679542](https://pubmed.ncbi.nlm.nih.gov/37679542/)
96. Hong L, Hu Z, Sun S, Tang X, Wang J, Tan Q, et al. Fast, sensitive detection of protein homologs using deep dense retrieval. *Nat Biotechnol.* 2024. <https://doi.org/10.1038/s41587-024-02353-6> PMID: [39123049](https://pubmed.ncbi.nlm.nih.gov/39123049/)
97. Xu Q, Tang C, Wang X, Sun S, Zhao J, Kang Z, et al. An effector protein of the wheat stripe rust fungus targets chloroplasts and suppresses chloroplast function. *Nat Commun.* 2019;10(1):5571. <https://doi.org/10.1038/s41467-019-13487-6> PMID: [31804478](https://pubmed.ncbi.nlm.nih.gov/31804478/)
98. Wang X, Zhai T, Zhang X, Tang C, Zhuang R, Zhao H, et al. Two stripe rust effectors impair wheat resistance by suppressing import of host Fe-S protein into chloroplasts. *Plant Physiol.* 2021;187(4):2530–43. <https://doi.org/10.1093/plphys/kiab434> PMID: [34890460](https://pubmed.ncbi.nlm.nih.gov/34890460/)

99. Bao X, Hu Y, Li Y, Chen X, Shang H, Hu X. The interaction of two *Puccinia striiformis* f. sp. tritici effectors modulates high-temperature seedling-plant resistance in wheat. *Mol Plant Pathol*. 2023;24(12):1522–34. <https://doi.org/10.1111/mpp.13390> PMID: [37786323](#)
100. Whisson SC, Boevink PC, Moleleki L, Avrova AO, Morales JG, Gilroy EM, et al. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature*. 2007;450(7166):115–8. <https://doi.org/10.1038/nature06203> PMID: [17914356](#)
101. Lévesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, Huitema E, et al. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol*. 2010;11(7):R73. <https://doi.org/10.1186/gb-2010-11-7-r73> PMID: [20626842](#)
102. 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;21(5):1573–91. <https://doi.org/10.1105/tpc.109.066324> PMID: [19454732](#)
103. Ridout CJ, Skamnioti P, Porritt O, Sacristan S, Jones JDG, Brown JKM. Multiple avirulence paralogues in cereal powdery mildew fungi may contribute to parasite fitness and defeat of plant resistance. *Plant Cell*. 2006;18(9):2402–14. <https://doi.org/10.1105/tpc.106.043307> PMID: [16905653](#)
104. Godfrey D, Böhlenius H, Pedersen C, Zhang Z, Emmersen J, Thordal-Christensen H. Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. *BMC Genomics*. 2010;11:317. <https://doi.org/10.1186/1471-2164-11-317> PMID: [20487537](#)
105. Catanzariti A-M, Dodds PN, Lawrence GJ, Ayliffe MA, Ellis JG. Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell*. 2006;18(1):243–56. <https://doi.org/10.1105/tpc.105.035980> PMID: [16326930](#)
106. Bordin N, Sillitoe I, Nallapareddy V, Rauer C, Lam SD, Waman VP, et al. AlphaFold2 reveals commonalities and novelties in protein structure space for 21 model organisms. *Commun Biol*. 2023;6(1):160. <https://doi.org/10.1038/s42003-023-04488-9> PMID: [36755055](#)
107. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, et al. TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. *Mol Plant*. 2023;16(11):1733–42. <https://doi.org/10.1016/j.molp.2023.09.010> PMID: [37740491](#)
108. DeLano W. The PyMOL molecular graphics system. *CCP4 Newsl Protein Crystallogr*. 2002;40:82–92.
109. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;42(Web Server issue):W320–4. <https://doi.org/10.1093/nar/gku316> PMID: [24753421](#)