



HAL
open science

Surface frustration re-patterning underlies the structural landscape and evolvability of fungal orphan candidate effectors

Mark Derbyshire, Sylvain Raffaele

► To cite this version:

Mark Derbyshire, Sylvain Raffaele. Surface frustration re-patterning underlies the structural landscape and evolvability of fungal orphan candidate effectors. *Nature Communications*, 2023, 14 (1), pp.5244. 10.1038/s41467-023-40949-9 . hal-04222363

HAL Id: hal-04222363

<https://hal.inrae.fr/hal-04222363v1>

Submitted on 29 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.





Distributed under a Creative Commons Attribution 4.0 International License

Surface frustration re-patterning underlies the structural landscape and evolvability of fungal orphan candidate effectors

Received: 10 January 2023

Accepted: 9 August 2023

Published online: 28 August 2023

 Check for updatesMark C. Derbyshire¹ & Sylvain Raffaele²  


Pathogens secrete effector proteins to subvert host physiology and cause disease. Effectors are engaged in a molecular arms race with the host resulting in conflicting evolutionary constraints to manipulate host cells without triggering immune responses. The molecular mechanisms allowing effectors to be at the same time robust and evolvable remain largely enigmatic. Here, we show that 62 conserved structure-related families encompass the majority of fungal orphan effector candidates in the Pezizomycotina subphylum. These effectors diversified through changes in patterns of thermodynamic frustration at surface residues. The underlying mutations tended to increase the robustness of the overall effector protein structure while switching potential binding interfaces. This mechanism could explain how conserved effector families maintained biological activity over long evolutionary timespans in different host environments and provides a model for the emergence of sequence-unrelated effector families with conserved structures.

Pathogens cause disease by manipulating host cell physiology to create a niche favorable to invasive growth and the completion of their life cycles. To do so, pathogens secrete small effector molecules active on host cells^{1,2}. Pathogen effectors fall into diverse molecular classes. Secondary metabolites such as fumonisins and AAL toxins are produced by the fungal pathogens *Fusarium* and *Alternaria* to disrupt host sphingolipid metabolism and trigger cell death³. The gray mould fungus *Botrytis cinerea* uses transposon-derived small RNAs to interfere with the expression of plant defense genes⁴. Several fungal toxins are cyclic peptides post-translationally modified, such as *Cochliobolus* victorin that requires a specific plant protein target to confer virulence⁵. Many effectors are small secreted proteins with very diverse activities, among which are suppressing host immunity, hijacking nutrients, masking the pathogen presence, interfering with host development, and killing host cells^{6–8}.

Some pathogen effectors are active in the host cell while others function in the intercellular space⁹. Their activity often involves molecular interactions with host target proteins. For instance, the small cysteine rich effector Avr2 of the fungus *Cladosporium fulvum* acts as a protease inhibitor targeting the tomato protease Pip1 to block

its activity, which contributes to plant defense¹⁰. Resistant host genotypes harbor receptor proteins triggering defense responses upon binding to pathogen effectors or detecting the product of their activity. In tomato, the receptor-like protein Cf-2 recognizes Avr2 bound to Rcr3, a decoy protease closely related to Pip1, resulting in hypersensitive response and resistance¹⁰. The molecular interactions drive strong selective constraints on effector genes to manipulate their host targets efficiently while evading immune recognition by the host.

Selective constraints on effector genes left signatures in the genome of eukaryotic pathogens^{11,12}. Typical effector genes evolve fast, showing frequent duplications, gain and loss, and a high degree of sequence divergence¹³. They often reside in repeat-rich genome compartments thought to accommodate high plasticity with limited deleterious fitness effects¹⁴. Fast effector gene evolution leads to expanded effector repertoires of hundreds of candidate effectors in the genomes of many filamentous microbes¹⁵. Effector candidates are often encoded by orphan genes that lack detectable sequence homologs outside their species or lineage, and therefore showing a patchy phylogenetic distribution. Effectors are typically small secreted proteins (SSPs) harboring an N-terminal secretion signal and several

¹Centre for Crop and Disease Management, School of Molecular and Life Sciences, Curtin University, Perth, Australia. ²Laboratoire des Interactions Plantes Micro-organismes Environnement (LIPME), INRAE, CNRS, Université de Toulouse, 31326 Castanet-Tolosan, France.  e-mail: sylvain.raffaele@inrae.fr

cysteine residues. In fungi, >95% of SSPs lack known functional domains¹⁶. This makes the identification of effector candidates, the inference of their evolutionary origin and their biological and molecular function very challenging.

Using knockouts in filamentous plant pathogens, several studies have provided evidence that individual effectors contribute to virulence. While the genome of the oomycete pathogen *Phytophthora infestans* harbors >500 effector genes of the RXLR type, silencing of the single Avr3a impairs virulence^{17,18}. Remarkably, several mutants of the smut fungus *Ustilago maydis* were strongly reduced in virulence and the corresponding effectors showed diverse activities in plant cells^{19,20}. More often, however, the deletion of single effectors leads to minor or no reduction in virulence²¹. The simultaneous deletion of multiple effectors is required to significantly reduce the virulence of the bacterial pathogen *Pseudomonas syringae* pv. *tomato*²² and the fungus *B. cinerea*²³, highlighting some apparent degree of redundancy in the repertoire of pathogen effectors. In *B. cinerea*, which infects hundreds of plant species, the effect of multiple knockout on virulence differed markedly according to host species and tissues inoculated. This supports the view that the activity of a set of effectors depends on the repertoire of their targets²³. Yet, the number of effector genes encoded by pathogen genomes does not increase with the range of hosts they infect¹⁶, suggesting that some effectors are active on multiple host species, possibly on multiple targets. One example is *P. infestans* Avr3a that contributes to virulence through its interaction with plant E3 ligase, GTPase dynamin-related protein, cinnamyl alcohol dehydrogenase and perhaps other nuclear proteins⁶.

Another layer of complexity emerges from interference between the activity of several effectors, either through epistatic interactions²⁴ or by cross-suppression of immune recognition^{25,26}. Finally, effectors may promote virulence not by directly targeting host processes but rather by altering the host microbiota^{27,28}. In summary, functional studies of effectors revealed that the simple equation “one effector, one biological activity” is generally false. Many pathogen effectors are promiscuous (active on multiple targets with no systematic fitness effect) or multifunctional (their activity on multiple targets confers an adaptive advantage).

Structural analysis revealed another level of redundancy and multifunctionality in the repertoire of filamentous pathogen effectors. The structure of several oomycete RXLR effectors uncovered a common four-helix bundle, folding around a WY domain²⁹. This fold was proposed to serve as a structural scaffold allowing robust function and surface plasticity, enabling WY domain effectors to bind to diverse host proteins³⁰. The structure of *Magnaporthe oryzae* AVR1-CO39 and AVR-Pia effectors exhibit a typical six-stranded β -sandwich fold analogous to *Pyrenophora tritici-repentis* ToxB necrotrophic effector. These effectors contribute to virulence on rice and wheat respectively and define the *Magnaporthe* Avrs and ToxB-like (MAX) fold³¹. Effectors from several *Blumeria* fungal species exhibit a structure analogous to RNase-Like Proteins associated with Haustoria (RALPH)³². *Fusarium oxysporum* f. sp. *lycopersici* (*Fol*) effector Avr2/SIX3 shows a seven-stranded β -sandwich fold analogous to *Pyrenophora tritici-repentis* ToxA³³. *Fol* Avr1/SIX4 and Avr3/SIX1 display a similar *Fol* dual-domain (FOLD) structure³⁴. The AvrLm3, AvrLm5-9 and AvrLm4-7 effectors of *Leptosphaeria maculans* share the *Leptosphaeria* Avirulence and Suppressing (LARS) fold with candidate effectors from 13 fungal species²⁶. Structural predictions used to identify candidate effectors in the secretome of fungal pathogens revealed additional protein folds conserved across species³⁵⁻³⁸.

Remote homology clustering allowed grouping 6538 fungal proteins related to effectors into 80 clusters spanning multiple pathogen species³⁹. The structural genomics analysis of the secretomes of 21 fungi identified 485 structure-related families of five or more proteins, 374 of which were previously unknown³⁸. Families containing known effectors had members from multiple pathogen and non-pathogen

species, suggesting that apparently novel effector groups could have originated from conserved fungal proteins through duplication and sequence divergence. Strikingly, none of the structure-related families had been detected by classical sequence similarity searches alone, indicating a high degree of sequence divergence and highlighting that their structures are robust to heavy mutational burden. The best characterized WY and MAX folds revealed that structural similarity does not imply effector functional similarity⁴⁰. *M. oryzae* MAX effector AVR-Pik binds and stabilizes rice susceptibility proteins with heavy metal-associated (HMA) domains⁴¹, *M. oryzae* MAX effector AvrPiz-t binds to a RING E3 ubiquitin ligase to suppress rice immunity⁴² and *P. tritici-repentis* MAX effector ToxB triggers host cell-death and susceptibility in wheat carrying sensitive *Tsc2* alleles⁴³. Conserved effector structures therefore enable a high degree of molecular promiscuity and evolvability. Robustness to mutation and the capacity to acquire new functions are not directly constrained by selection for protein function, and vary among proteins with nearly identical structure⁴⁴. Whether a specific evolutionary trajectory promotes the acquisition of robustness and evolvability in pathogen effectors is enigmatic.

Orphans are a particular class of protein for which no sequence homology with characterized proteins has been detected, indicative of extreme sequence divergence or recent gene birth. The extent to which orphans in fungal secretomes adopt similar protein structures and the evolutionary properties of such structures remain elusive. In this work, we deployed computational structural phylogenomics on ~5,000 secreted proteins from 20 fungal species and structural similarity network analysis to reveal the structural landscape of orphan candidate effectors (OCEs) at the entire subphylum level. Sixty-two structure-related families aggregated a majority of OCEs and were widely represented among species with diverse lifestyles and host ranges. Subsets of OCE families showed deep homology indicative of a distant common ancestry, with sequence and structural variation targeting mostly surface-exposed residues. Using in silico mutation scans, we identified clusters of co-evolving surface residues with contrasting effects on the overall fold robustness. Ancestral structure reconstruction revealed that fluctuations in local frustration of surface residue during OCE evolution associate with mutational robustness of conserved OCE families. We propose that re-patterning of surface frustration promotes the emergence of effector families that are both robust and highly evolvable.

Results

Ancient generic folds dominate the structural landscape of orphan candidate effectors

To explore the structural landscape of fungal orphan candidate effectors (OCEs), we predicted the complete repertoire of 3D structures for OCEs in twenty fungal species covering the phylogenetic and lifestyle diversity of the Pezizomycotina (Fig. 1A, B). Starting from complete genome sequences and the corresponding predicted proteomes, our pipeline selected proteins (i) including a predicted secretion signal, (ii) of mature size shorter than 300 amino-acids, (iii) lacking recognized PFAM protein domains, and (iv) with intrinsic disordered regions covering less than 50% of the sequence (Fig. 1A, B). From an initial set of 227,823 predicted proteins, we selected 4987 Orphan Candidate Effectors (OCE) for structure prediction using AlphaFold⁴⁵. We obtained 3927 OCE structures of average pLDDT score higher than 50. In subsequent analyses, we focused on the best resolved structures to reach a mean pLDDT > 80 for 99.67% models reported in this work (Supplementary Fig. 1).

To classify orphan effector predicted structures according to similarity, we performed pairwise comparisons with 3911 OCE structures in DALI⁴⁶. 477 proteins (12.2%) had no significant structural similarity ($Z < 2$) to any other OCE in the dataset. For a more conservative estimate, we focused on the 2561 orphan effectors with at least three analogs at $Z \geq 5.2$, leaving out 1350 (34.5%) singletons. With

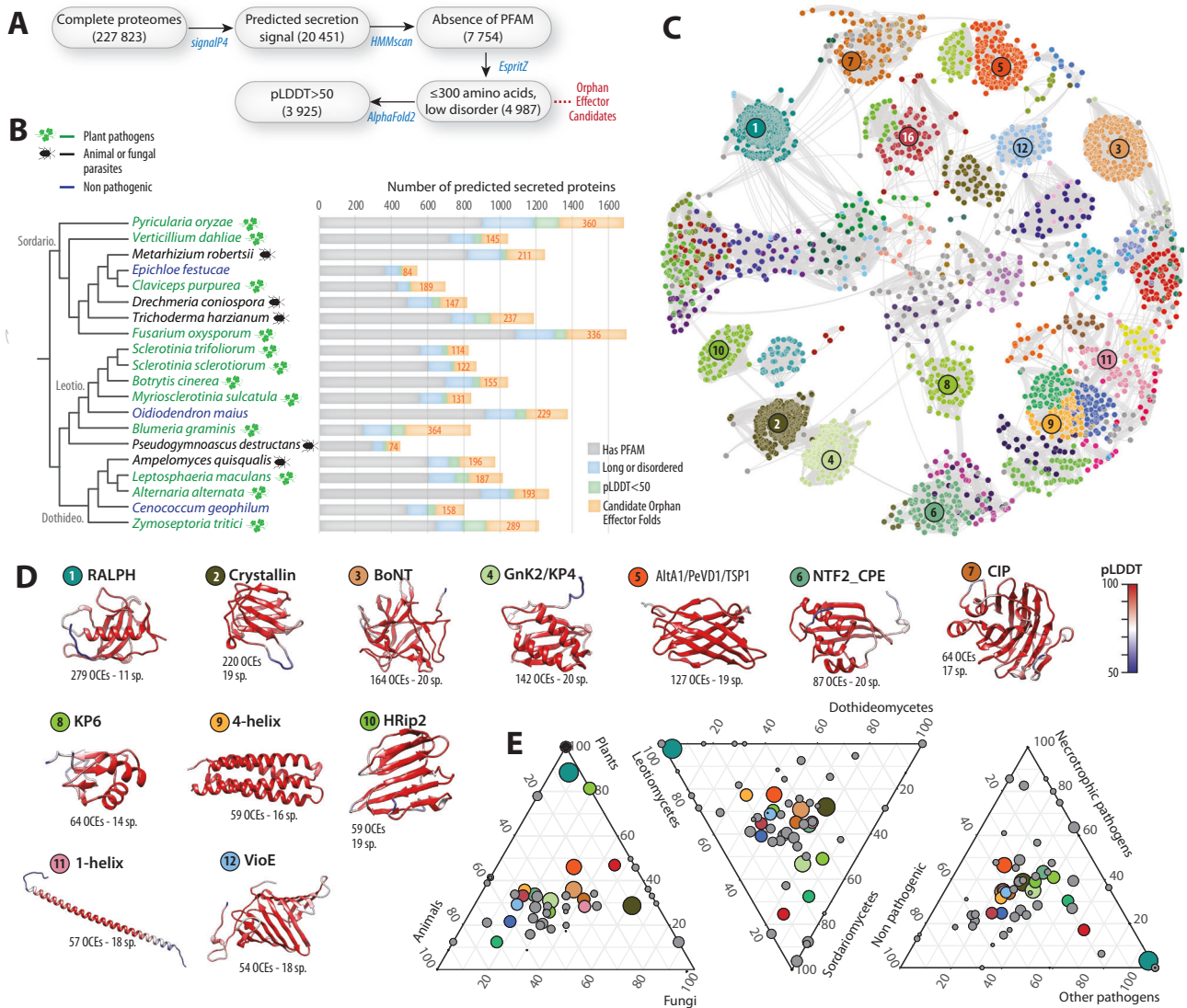


Fig. 1 | Structural relationship between orphan candidate effectors in 20 fungal genomes. **A** Bioinformatics pipeline for the systematic analysis of structures of orphan candidate effectors (OCEs) in fungal genomes. Filtering and analysis steps are shown as gray boxes with the number of retained proteins indicated between brackets. Analytical tools are indicated in blue. Out of 227,823 predicted proteins, 20,451 had a secretion signal, 4987 we selected for structure prediction, 3925 had a reliable structure predicted. **B** Fungal species selected to cover the phylogenetic and lifestyle diversity of the Pezizomycotina. For each proteome, the distribution of secreted proteins with different characteristics based on the pipeline is indicated as a histogram. **C** OCE structural similarity network with vertices colored according to

detected communities forming structure-related families. The 12 major families including ≥ 54 proteins are numbered by decreasing protein count. **D** Representative predicted structures corresponding to the 12 major families. Numbered and colored stickers refer to communities shown in **(C)**. Residues are colored according to their pLDDT score. The number of OCEs and species in which they occur (sp.) is indicated. **E** Contribution of fungal species classified by lineage, host type and lifestyle to each of the 62 major families. Each circle represents one OCE fold, sized according to the number of proteins it contains. The 12 major families are colored as in **(D)**. Source data are provided as a Source data file.

those, we built a structural similarity network featuring OCE predicted structures as vertices and DALI Z-scores as weighted edges (Fig. 1C). The structural similarity network featured nodes of high connectivity, corresponding to protein folds shared between multiple orphan effectors. In agreement, the network has a high clustering coefficient (0.86) and high heterogeneity (0.579) but low density (0.195) and low centralization (0.35).

Six effectors from the network had experimentally determined structures (Supplementary Fig. 2). Consistent with previous reports³⁸, the match between experimentally determined structures and AlphaFold2 predictions was very good, with RMSD ranging from 0.468 to 3.94 angstroms. To characterize OCE structure-related families, we performed community detection on the OCE structure similarity network and compared OCE predicted structures to those in the PDB database. Combining hierarchical community detection with HiDef

and manual curation, we identified 62 families of at least five effector candidates, representing 2371 OCEs across the twenty fungal genomes (60.4% of all OCE structures predicted with pLDDT > 50, Fig. 1C). The most abundant was the RALPH family found in 279 OCEs from 11 species, among which 262 were from the powdery mildew pathogen *Blumeria graminis* f. sp. *hordei* (Fig. 1C, D). The most widespread families, retrieved in all 20 species, were related to toxin folds: the β -trefoil found in *Clostridium* neurotoxins and plant ricins, the β - $(\alpha\beta)_2$ sandwich of KP4 fungal killer toxin and plant ginkbilobin-2, the $\alpha\beta_4$ core sandwich of the nuclear transport factor 2 and *Clostridium perfringens* enterotoxin. Eighteen families were detected in over 80% of the species (Fig. 1C, D, Supplementary Figs. 3, 4), including additional toxin folds such as the β_8 double Greek key of $\beta\gamma$ -crystallin and the yeast killer toxin WmKT⁴⁷, and the $(\alpha\beta\alpha)_2$ sandwich of the *Ustilago maydis* KP6 killer toxin⁴⁸. Among the eighteen broadly conserved

families, we identified five previously associated with known fungal effectors (Fig. 1D): RALPH⁴⁹, Alt-A1^{50–52}, HRip2⁵³, and the 5-helix bundle found in *Drechmeria coniospora* g7941 effector (6zpp). We identified three additional known effector families (Fig. 1D): MAX (44 OCEs in 3 species)³¹, Avr2/ToxA (39 OCEs in 6 species)³³ and Tox3 (6 OCEs in 5 species)³⁴.

Out of 62 major OCE families, only three were specific to a single fungal species, suggesting that specialization in these OCEs is limited. To support this claim, we analyzed the distribution of the 62 families across fungal lineages, host group and lifestyle (Fig. 1E). Forty four OCE families were detected in Leotiomycetes, Sordariomycetes and Dothideomycetes with cases of lineage-specific enrichments (e.g MAX and Tox3 folds enriched in the Sordariomycetes). Only five OCE families were restricted to a single lineage. Thirty-five OCE families were detected in pathogens infecting plant, fungal and animal hosts, while 13 OCE families were restricted to fungi infecting only one of these host groups. Twelve families were restricted to plant pathogens including MAX and SIX3. Forty-six OCE families were detected in pathogens with a mutualistic, a necrotrophic and other pathogenic lifestyles. Apart from the KP6 family, which was absent from animal pathogens, all the most abundant OCE families were detected in fungi from all lineages, infecting all host groups and with all lifestyles.

In agreement with previous analyses of fungal secretomes^{38,39}, we highlight a few OCE families highly represented at the subphylum level, consistent with deep common ancestry. Although considered as orphans based on classical sequence homology searches, many OCEs adopt effector and toxin folds described previously and distributed broadly across fungal species with various affinities and lifestyles. This suggests that ancient OCE folds remained adaptive in diverse biological contexts, and in spite of considerable sequence divergence.

OCE families display a stable core with flexible surface-exposed regions

The detection of sequence-unrelated OCEs with similar 3D structures suggests that modern OCEs derived from ancestral structures which remained stable over evolution, in spite of extensive sequence divergence^{31,38}. To document the extant diversity in major OCE families, we first compared sequence-based evolutionary distance to structural similarity (Fig. 2A, Supplementary Fig. 5). We performed the same analysis for Histone 2A from the twenty fungal species used in our analysis to serve as a control. For all protein groups, structural similarity decreased regularly for evolutionary distance 0 to 200, and tended to stabilize at Z-score -5.0 beyond this distance. The number of proteins showing unexpectedly high structural similarity to other members of the family (above 97.5% prediction interval) was 0 for Histone 2A, 68 for Alt-A1, and 89 for KP4 toxins. Among OCE families, the proportion of proteins with unexpectedly high structural similarity to other members ranged from 80.4% (Crystallin) to 23.4% (KP6). OCE families therefore maintain a high structural similarity in spite of high evolutionary distance.

In some effector families, variable residues tend to reside at the surface of proteins³¹. To test whether structural similarity related to amino acid burial in OCEs, we analyzed structure-based alignments of OCEs representative of sequence diversity in Alt-A1 and BoNT families (Fig. 2B–D). Although some of the selected OCEs shared as little as 3% identity with each other, their structures superimposed with good accuracy. Residues with the highest relative surface exposure localized to regions where the structural alignments decreased in quality (Fig. 2C, Supplementary Movies 1, 2). We observed a global correlation between relative surface exposure and alignment deviation (RMSD, in Angstroms) for aligned residues, supporting a link between conformational flexibility and surface exposure (Fig. 2D). We also noted a tendency for more conserved residues to reside in buried, structurally stable regions, although the high sequence divergence among the

selected representative sequences did not allow to assess this trend accurately (Fig. 2D).

To analyze evolutionary relationships within OCE families, we used sensitive all-vs-all profile hidden Markov model (HMM) comparisons for all fungal BLASTp homologs in the NCBI nr database of 11 of the major OCE families (Fig. 2E, Supplementary Fig. 6). In the Alt-A1 family, we clustered 11,251 amino acid sequences into 494 groups of proteins with >30% AA identity, and further grouped these into 13 super-clusters based on HMM comparisons. OCEs from our 20-genomes analysis appeared in all 13 super-clusters. Super-clusters 1 to 7 showed weak homology relationships, supporting the view that ancient and broadly conserved protein folds were recruited as effectors in fungi with diverse lifestyles. Though there was structural similarity between OCEs and metazoan and bacterial proteins, we did not detect evidence of common ancestry between super-clusters 1 to 7 and 8 to 13. Therefore, although a single ancestral group dominates the structural family, we cannot rule out the possibility of convergence in structure of unrelated sequences. Similar conclusions could be drawn from analysis of the other 10 structural families, supporting the existence of OCE families identified based on structural similarity analyses. For eight of the 11 OCE families, at least 75% of the predicted effectors had remote sequence homology to structural hits from fungi in PDB (Supplementary Fig. 6). Over all, for homologs of the Alt-A1, BoNT, Gink2/KP4, HRip2, KP6, and NFT2/CPE OCE families, the largest super-clusters comprised 89%, 94%, 91%, 97%, 98%, 69%, and 89%, respectively, of HMMs and singletons (Supplementary Fig. 7). Whereas HMMs comprising homologs of OCEs from the atg, CIP, g7941, SIX3, and VioE families were spread more evenly across super-clusters, with the largest super-clusters comprising from 35% (SIX3) to 47% (g7941) of HMMs and singletons.

We conclude that major OCEs likely derive from ancient protein folds that accumulated surface variation and maintained a stable core despite extensive sequence divergence.

Modern OCEs accumulated co-selected mutations reducing structural divergence

To test the impact of mutations on major OCE families, we reconstructed ancestral OCE structures for clades of KP6 and Alt-A1 families (Fig. 3A). To highlight structural regions prone to sequence variation, we mapped the positions of modern variants from each clade on the ancestral protein structure (Fig. 3B, Supplementary Movies 3, 4). Unexpectedly, multiple weakly conserved residues occupied the core of the ancestral OCE structure, and amino acid conservation did not clearly discriminate between core and surface regions. To determine the extent to which these amino acid changes were selected independently, we calculated the frequency of mutation co-occurrence in modern OCEs. We shuffled extant variable residues 10,000 times to estimate a null probability of mutation co-occurrence and estimate co-selection *p*-values for observed mutation pairs. In the KP6 clade, 64 residues (84.2% of variable positions) had co-selection *p*-val <1.0E-10. In the Alt-A1 clade, 60 residues (38.7% of variable positions) had co-selection *p*-val <1.0E-10. Significant co-mutation relationships defined 6 groups of co-selected mutations both in KP6 and Alt-A1 (Fig. 3C, Supplementary Movies 5, 6). Apart from co-selected cluster 2 in KP6 (red, Fig. 3C), co-selected amino acids tended to distribute across the whole OCE structure rather than localize to specific sites of the protein.

To assess the effect of extant sequence diversity on OCE structural diversity, we compared structures predicted for extant OCEs with the structure of their common ancestor. The self-alignment of the ancestral structure served as a reference to estimate the structural divergence (SD) induced by natural diversity. In the KP6 clade, most extant variants had SD <2.0, indicative of a very mild change in structure, while in the Alt-A1 clade, median SD was 4.3 (Fig. 3D). To estimate the relative contribution of individual mutations to OCE structural

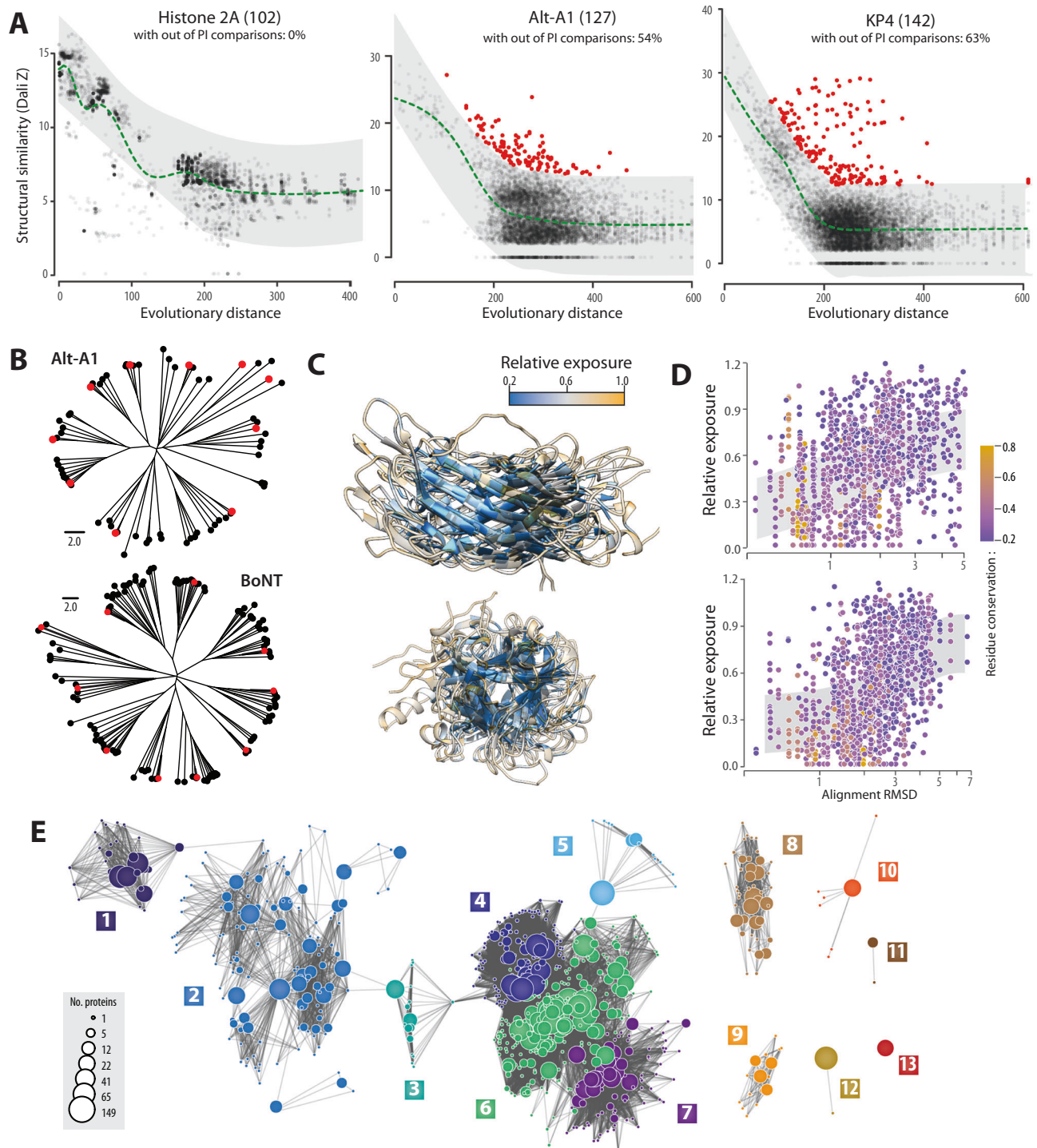
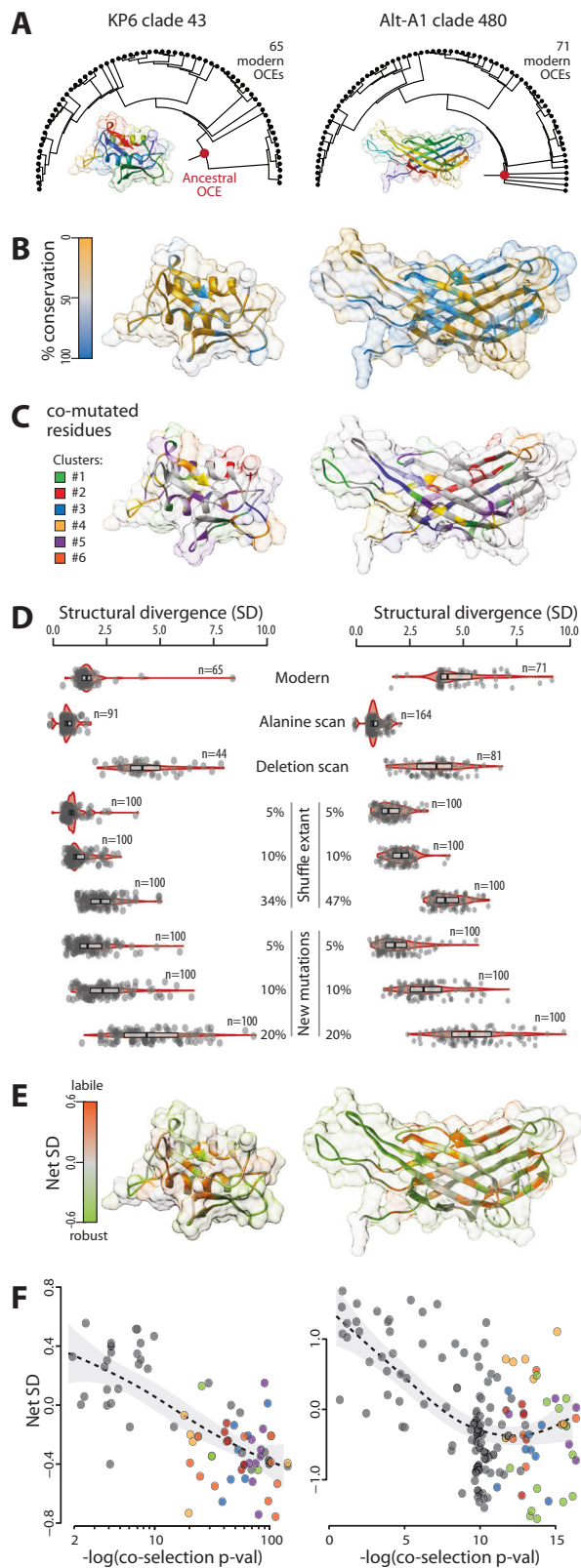


Fig. 2 | The extant sequence and structural diversity in major OCE families reveal stable cores with flexible surface-exposed regions. **A** Structural similarity (Dali Z-score) according to evolutionary distance (Jukes-Cantor corrected distance calculated on curated sequence alignments) for histone 2A and the OCE families Alt-A1 and KP4 toxins. The green dotted lines are the local average. Pairwise structure comparisons falling outside of the 97.5% prediction interval (PI, gray ribbons) are shown as red dots. Numbers in brackets indicate the number of protein structures compared. The percent of proteins with Z values above the predicted interval is indicated for each family. **B** Structure similarity neighbour-joining trees for members of the Alt-A1 and BoNT families, with representative structures

shown as red circles. **C** Structural alignment of the corresponding representative proteins colored according to the residue relative exposure. **D** Relative residue exposure according to local structural alignment accuracy (RMSD, in Angstrom). Gray ribbons show LOESS regression-centered 50% prediction intervals. Dots are colored according to residue conservation. **E** The overall diversity of the Alt-A1 family obtained through sensitive HMM comparisons of homologs from the NCBI nr database. Each circle corresponds to a cluster of 1 to 169 sequences, with homology relationships shown as edges. The 13 superclusters are numbered and shown in different colors. Superclusters 1 to 7 are connected by homology to a single member. Source data are provided as a Source data file.



divergence and document the theoretical sequence diversity that the major OCE folds can withstand, we performed mutation scans followed by structure prediction and comparison on the KP6 and Alt-A1 ancestral proteins. We mutated individual amino acids to Alanine and used a sliding window deletion scan to estimate SD for individual residues in the ancestral KP6 and Alt-A1 proteins. Next, we shuffled

Fig. 3 | Mapping of robustness-increasing mutations based on natural diversity and in silico mutation scans in clades of KP6 and Alt-A1 OCEs. **A** Reconstruction of OCE ancestral structure in KP6 clade 43 and Alt-A1 clade 480. The phylogenetic trees used for ancestral sequence reconstruction show modern OCEs as black circles and ancestral nodes used in structure prediction as red circles. The AlphaFold models of the common ancestors are shown in rainbow colors. **B** Sequence conservation (%) in modern OCEs mapped onto the protein structure of their common ancestor. **C** Mapping of clusters of residues significantly co-mutated onto ancestral protein structures. **D** Structural divergence (SD) caused by extant natural variants ('Modern') and artificial mutations in ancestral OCEs. Structural divergence between variant OCEs is expressed relative to ancestral OCE self-alignment score. Boxplots show median values (thick line), first and third quartile values (box) and 1.5 times the interquartile range (whiskers). **E** Mapping of residues net SD effect onto the structure of the ancestral OCE. Net SD corresponds to the difference between SD obtained after the mutation/deletion of individual residues and SD obtained after multiple simultaneous mutations. **F** Relationship between residues SD effect (Y-axis) and their probability of co-selection (X-axis, as \log_{10} of p -value). Residues are colored as in (C) according to the cluster they belong to. The dotted line shows local average, the gray area corresponds to the 95% confidence interval. Source data are provided as a Source data file.

existing variants and introduced multiple new mutations randomly to modify simultaneously 5% to 47% of the ancestral sequence and assess the SD effect of multiple mutations on the ancestral protein structure (Fig. 3D). All Alt-A1 and 97% of KP6 multiple mutants had a SD smaller than the sum of SD for individual mutated residues they contain, indicating that the effect of individual mutations on SD is not strictly additive. We hypothesized that the missing SD in multiple mutants was due to compensatory effects contributing to the robustness of OCE folds. We used this information to derive the net SD effect of mutations at each position (see methods). This approach identified 47 residues (51%) in KP6 and 92 residues (56%) in Alt-A1 with negative net SD effect that could increase the mutational robustness of OCEs. Robustness-increasing residues tended to reside away from the OCE structural core (Fig. 3E, Supplementary Movies 7, 8).

In our hypothesis, mutations at robustness-increasing positions would compensate for structural variation due to mutations occurring elsewhere in the protein, maintaining a conserved structure despite sequence divergence. To support this model, we analyzed the SD effect according to the likelihood of co-selection in KP6 and Alt-A1 clades (Fig. 3F). On average, the SD effect decreased with probability of co-selection in KP6 and in Alt-A1. Within clusters of co-selected residues, the amplitude of the net SD effect covered 17 to 49% of the total amplitude in the corresponding protein, supporting the view that robustness-increasing residues co-evolved with residues with a strong SD effect.

OCEs evolved through oscillations in surface residue frustration

Proteins acquiring mutations enhancing stability are more likely to tolerate destabilizing mutations, which may confer functional benefits^{55,56}. Indeed, surface residues showing high frustration, corresponding to unfavorable interactions in the protein native state, often contribute to binding interactions releasing frustration⁵⁷. To test whether OCEs exhibit a frustrated and evolvable surface surrounding a thermodynamically stable core, we assessed the relationship between frustration index⁵⁸ and relative surface exposure. Across the three families we tested (AltA1, KP6, and Gink2/KP4), frustration index decreased (corresponding to higher levels of frustration) with residue surface exposure (Spearman's $\rho = -0.4$, Fig. 4A). Highly frustrated amino acids had a significantly higher average relative surface exposure of 0.66 compared with 0.52 for minimally and neutrally frustrated residues ($P = 2.2e-16$, Fig. 4B, Supplementary Fig. 8). To identify OCE regions prone to frustration variation, we calculated frustration index variance in four clades of four different families, and mapped these

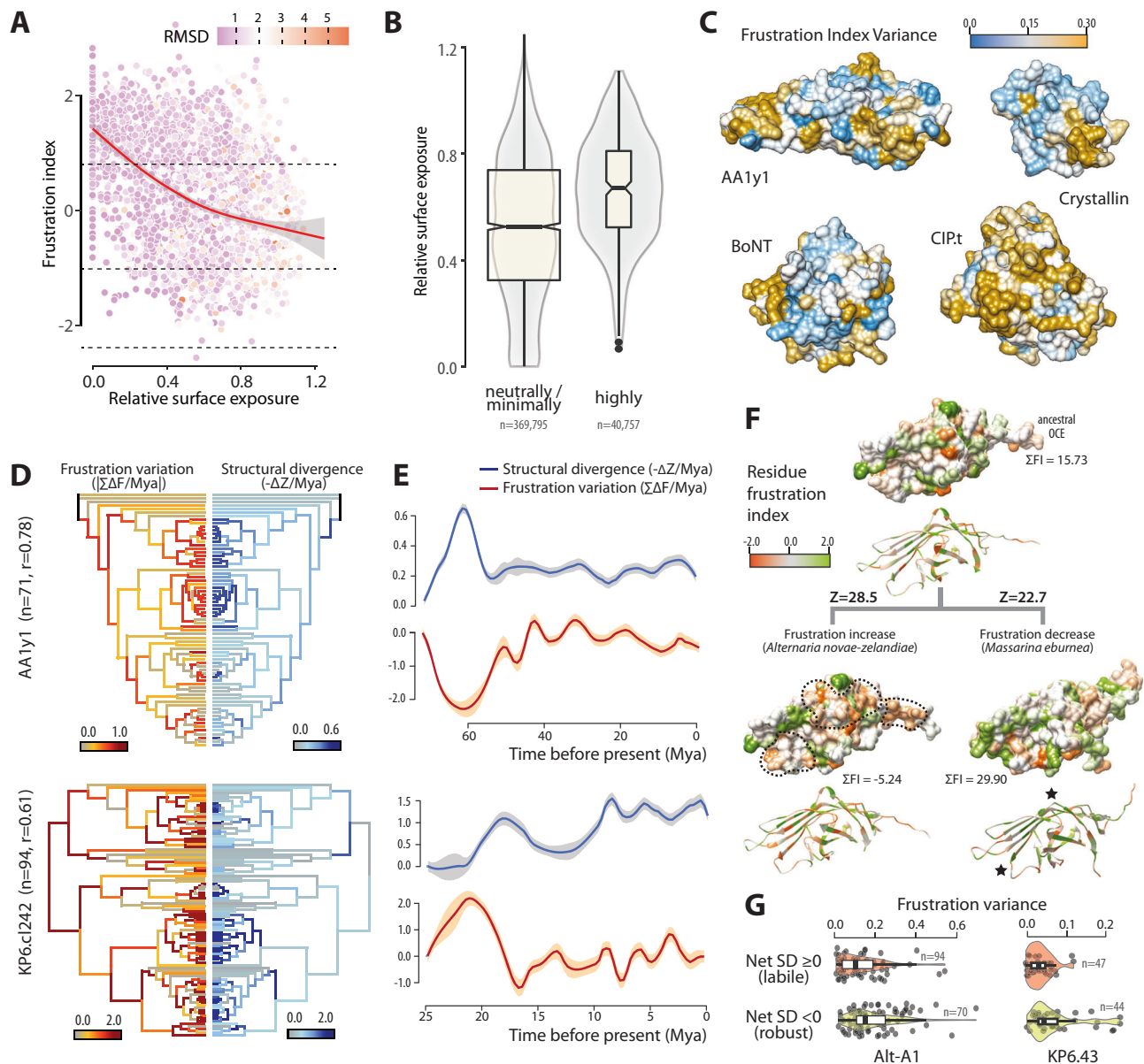


Fig. 4 | Re-patterning of surface residue frustration buffers structural variation in OCEs. **A** Frustration index negatively correlates with relative surface exposure in Alt-A1, KP6, and GNK2/KP4 OCE families. Frustration index decreases with the degree of residue frustration. The red line shows loess regression, gray area is the 90% confidence interval, dotted lines delimit minimally (0.78) and highly (−1.0) frustrated thresholds. Data points are colored according to C α structural deviation from predicted ancestral structure (RMSD, root mean square deviation, in Angstroms). **B** Comparison of relative surface exposure for residue classified as neutrally/minimally frustrated and highly frustrated in Alt-A1, KP6, and GNK2/KP4 OCE families. Boxplots show median values (thick line), first and third quartile values (box) and 1.5 times the interquartile range (whiskers). **C** Frustration index variance in clades from four OCE families mapped on the structure of the reconstructed clade ancestor. **D** Frustration variation (red, left side) and structure divergence (SD, blue, right side) rates during the evolution of OCEs from an Alt-A1 and a KP6 clade mapped on time-calibrated phylogenies. *N* is the number of modern OCEs per

group, *r* is Pearson's product-moment correlation between frustration variation and structural divergence across all branches of the tree. **E** Frustration variation (red) and structural divergence (blue) over time across every possible evolutionary path in trees shown in (D). Lines show mean value from a loess regression, shaded areas are the 95% confidence interval. **F** Reconstructed evolution of structure and residue frustration in two lineages of members of the Alt-A1 family. Each protein structure is shown as molecular surface (top) and ribbon diagram (bottom), with the reconstructed ancestral protein at the top of the panel and two of its modern descendants at the bottom. *Z*, structural similarity score; Σ FI, sum of frustration index for all residues in the protein. Black dotted lines delimit patches of increased frustration, black stars indicate the position of highly variable loops. **G** Frustration variance for residues showing positive (green) and negative (red) SD effect in mutational scans of two OCE groups. Mya, Million years ago. Boxplots show median values (thick line), first and third quartile values (box) and 1.5 times the interquartile range (whiskers). Source data are provided as a Source data file.

values on protein structures (Fig. 4C). The four families showed multiple surface patches with high frustration variance that could mediate differential binding with diverse molecular partners.

To document the evolution of structure and frustration in major OCE families, we built maximum likelihood phylogenies, reconstructed ancestral protein structures, and analyzed frustration for 15 HMM sequence clusters covering six OCE families and representing

1307 modern OCEs and 1289 ancestral OCEs. Using time-calibrated trees, we estimated the rate of SD and amino-acid frustration along each branch of the phylogenies (Fig. 4D). We observed a general correlation between the rate of frustration variation and the rate of SD (median Pearson's $r = 0.66$) along branches, indicating that amino-acid thermodynamic frustration is a major driver of structural change in OCEs (Supplementary Figs. 9, 10). The path from root to tip passed

through several frustration and SD local maxima. To document the evolutionary dynamics of frustration and SD in selected OCE clades, we collected these metrics along every evolutionary path in trees, using estimated branching times as a clock (Fig. 4E). It revealed cycles of frustration maxima followed by frustration release. Frustration decrease roughly coincided with bursts of SD relative to previous nodes, indicating that the release of frustrated contacts is a major step in the structural evolution of OCEs. In some OCE clades, modern proteins close to their frustration maximum showed reduced structural divergence from the clade ancestor compared to modern proteins with low frustration (Fig. 4F). In agreement, residues with a robustness-increasing effect (net SD < 0) identified in Alt-A1 and KP6 clusters (Fig. 3E) respectively showed 1.5 and 2.0 times higher frustration variance than positions with a destabilizing effect (t test p -val = 1E-03 and 8E-04, respectively, Fig. 4G). These results suggest that mutations increasing OCE frustration contribute to sequence and functional diversity while maintaining the overall protein structure robust, leading to the emergence of large families of sequence unrelated, structurally conserved effector candidates.

Discussion

Pathogen effectors are engaged in a molecular arms race with the host resulting in conflicting evolutionary constraints to subvert host physiology efficiently on the one hand, and avoid detection by the host immune system on the other hand. In this work, we show that conserved fungal effectors diversified through re-patterning of frustrated surface residues. The underlying mutations tend to increase the robustness of the overall effector protein structure while switching potential binding interfaces. This mechanism could explain how conserved effector families have maintained biological activity over long evolutionary timespans and in different host environments, and diversified exposed residues to escape host immune detection. The finding that effector proteins can withstand high mutational burden by building up surface frustration provides a model for the emergence of widely distributed families of sequence-unrelated effector families with analogous structures.

The limited structural information available to date suggests that some oomycete effector families adopt a conserved modular core with variable surface residues to accommodate evolutionary constraints^{11,29,30}. Low sequence homology and the absence of distinctive motifs have prevented the large-scale detection of effector families in fungi. Recent structural studies and the advent of computational structural genomics have identified sequence-unrelated families of fungal effectors^{29,31,36}. Focusing specifically on small orphan secreted proteins, we reveal that the majority of fungal secreted proteins with no recognizable domains, the secretome dark matter, is made of families of structurally analogous effector candidates. Most of these OCE families are widely conserved across fungi with diverse lifestyles. In these families, amino-acid and structure conservation decreased with relative surface exposure, generalizing the model of effectors having a stable core and a highly variable surface, and expanding the known repertoire of effector stable cores that tolerate extreme surface variation.

We focused on 62 effector families that make up the majority of the Ascomycete OCE repertoire. For the most widespread families, sensitive HMM searches identified distant homologs in hundreds of species^{39,59}, suggesting that effector families may have deep common ancestry. A few OCE families were lineage-specific or showed species-specific expansions. If OCE families were often represented by several members in each genome, massive expansions like in the case of oomycete RXLR and CRN effectors¹⁷ were rather rare (e.g., RALPH-like in *B. graminis*, MAX in *P. oryzae* and *C. purpurea*), suggesting that segmental gene duplication may not be a major contributor to the evolution of fungal OCE families. Two major OCE families (GNK2/KP4 and KP6) showed structural similarity to yeast and filamentous fungi

killer toxins⁶⁰. In strains of the smut fungus *Ustilago maydis*, KP4 and 6 are encoded by persistent dsRNA viruses, and secreted by the fungus to kill competing strains that do not contain cognate toxin resistance genes. The *U. maydis* killer phenotype is inherited by horizontal gene transfer (HGT)⁶¹, and HGT to plants has been documented for KP4⁶². Furthermore, HGT is a key process in the inheritance of OCEs from the SIX and ToxA families^{63,64}, and in the emergence of pathogenicity in *Metarhizium* entomopathogens⁶⁵, suggesting that HGT may be an important process in the evolution of OCEs.

Based on the panel of Ascomycete species we selected, we estimate that ~70 families of structure-related OCEs cover the majority of the fungal uncharacterized effectome (Supplementary Fig. 11). Grouping into structural families can guide the functional characterization of pathogen effectors. A complete understanding of fungal pathogenicity will require expanding the current effort to include (i) other filamentous pathogen lineages such as Basidiomycete fungi and oomycetes (see, ref. 38), (ii) OCEs that were excluded from our pipeline such as those longer than 300 amino-acids, harboring PFAM domains or with no canonical secretion signal, (iii) OCEs with no clear analogs in protein structure databases and that did not group into structure-related families.

The wide distribution of many OCE families suggests that some could have generic functions in pathogenicity, such as broad host toxicity^{66,67} or competition with other microbes^{28,68}. In spite of sharing a similar structure, sequence conservation in OCE families is generally very low, which may be the result of evolution to escape host or competing microbe recognition. It also raises the possibility that OCEs from the same family may have different biological activities. This view is notably supported by differential phytotoxic activity in families of Necrosis- and ethylene-inducing peptide 1-like proteins^{67,69} and other cell death-inducing proteins⁷⁰. Our results suggest that mutation-driven increase in surface frustration promotes new molecular interactions and neofunctionalization without major alteration to the ancestral fold. Studies on intrinsically disordered proteins showed that frustration promotes promiscuous molecular interactions, leading to versatile protein complex assembly, instead of specific binding^{71,72}. This property may be mediated by increased flexibility of surface loops⁷³. Increase of frustration in OCEs may therefore be a mechanism to expand the repertoire of their binding targets, contributing to the emergence of general-purpose effectors. This study raises hypotheses regarding the evolution of effectors function that remain to be tested experimentally. A major objective for future work will include determining whether the energy landscape of effector proteins associates with specific binding activity and biological function. Molecular resurrection experiments will be required to test the evolutionary scenarios emerging from our analyses. The Alt-A1 OCE family, with members characterized from *B. cinerea*, *S. sclerotiorum*, *T. virens*, and *V. dahliae*⁷⁴ is an attractive target for such experiments.

The intramolecular interaction network of proteins allows them to accommodate substitutions to a certain extent, while still maintaining their native fold⁷⁵. In large populations, as for many fungi and other microbes, proteins evolve greater stability, a biophysical property that is known to enhance both mutational robustness and evolvability⁴⁴. Experimental evolution approaches indeed showed that the emergence of new functions was facilitated by the use of a more robust evolutionary starting point⁷². Our analysis of natural diversity, in silico mutation scan, and ancestral structure reconstruction indicate that several OCE folds are highly robust to mutations, notably through the ability to accumulate surface frustration. This property could therefore facilitate functional switches in OCEs.

The analysis of evolutionary paths in two OCE clades indicated that frustration tends to fluctuate over time. This suggests that the amount of frustration that OCE scaffolds can withstand is limited. Following the principle of minimal frustration, we would expect highly

frustrated OCEs to unfold more rapidly⁷⁶, and thus to be active for shorter time periods or in a more restricted range of conditions. Alternatively, highly frustrated OCEs may fall in kinetic traps leading to less efficient folding, or show an increased risk of forming deleterious aggregates⁷⁷. Release of frustration during OCE evolution roughly corresponded to phases of structural divergence, creating cycles of frustration increase followed by structural re-arrangements. The tradeoff between structural stability and biological activity has been reported in diverse protein types and is a major bottleneck in protein engineering⁷⁸. The evolution of fungal OCEs recapitulates the three solutions typically employed in protein engineering: (i) OCE families evolve from highly stable parental proteins with a robust core, (ii) structural destabilization during functional evolution is minimized by increasing surface frustration and the co-selection of residues with contrasted stabilizing effects, (iii) OCEs evolutionary paths establish a long-term balance between structure and thermodynamic stability. Patterns of thermodynamic stability loss and restoration were reported during the evolution of new enzyme activities⁵⁵, suggesting that fluctuations in surface frustration may correspond to functional switches in OCEs. These frustration patterns imply that the structural OCE template is not lost over evolution, including following neofunctionalization, and remains available for further diversification. It may also revert to a previous native state leading to apparent structural convergence.

Progress in effector structural biology and structural genomics enabled by deep learning algorithms accelerated the pace of the characterization of filamentous pathogen effectors. We propose here an innovative structural and phylogenomics pipeline combining ancestral structure reconstruction, mutation scans, and frustration analysis to investigate the molecular mechanisms and evolution of effector proteins. The structural phylogenomics strategy applied to OCE families of interest should help identifying residues involved in protein-protein interaction, infer probable functional switches during OCE evolution, infer thermodynamic properties of effectors and their evolutionary potential. Ongoing progress in the functional characterization of fungal secreted proteins will allow formally testing and refining mechanistic and evolutionary models derived from structural phylogenomics studies.

Methods

Identification of orphan candidate effectors and structure prediction

Complete predicted proteomes for 20 Ascomycete fungi were downloaded from the public repositories listed in Source Data. Secreted proteins were predicted using signalP-4.1g for Cygwin⁷⁹ which is recommended for pathogen effector predictions⁸⁰. Signal peptides were trimmed and proteins longer than 300 amino acids were excluded from further analysis. Mature sequences of secreted proteins were searched for conserved domains with the hmmscan.pl script using the Pfam-A 35.0 database (dated 2021-11). Proteins with no Pfam-A hit of p -value < 0.01 were considered as orphans. Orphans were searched for intrinsic disorder regions using the EspritZ server version 1.3⁸¹, with prediction type X-ray and 5%FPR decision threshold. Proteins with intrinsic disorder regions spanning $>50\%$ of their mature sequence were excluded from further analysis, leaving 4987 mature sequences as input for structure prediction. We chose AlphaFold2 to model OCE 3D structures since it has been rigorously tested and shown to offer currently the best performances for effectors of filamentous plant pathogens³⁸. OCE structures were predicted using the ColabFold: AlphaFold2 w/MMseqs2 BATCH notebook with parameters `msa_mode: MMseqs2 (UniRef+Environmental)`, `num_models: 4`, `num_recycles: 3`, `stop_at_score: 100`. Structures with average pLDDT score < 50 were excluded from further analysis. Major OCE folds in our structural similarity network had an average pLDDT from 80.01 to 89.06, 99.67% of all OCE structures analyzed in this work had pLDDT > 80 .

Structural relationships analysis and rendering

Structural similarity between the 3925 OCE structures was calculated using the all-versus-all function in DALI Lite 5.0, which showed top performances with the alignment of small protein domains as OCEs⁸². The resulting matrix of pairwise similarity Z-scores was converted into a structural similarity network with the igraph package in RStudio v1.4.1106, filtering out edges of weight (Z-score) < 5.2 and nodes with degree < 3 and using the Kamada-Kawai force-directed layout algorithm. The resulting network was exported as a.gml file using the plotCytoscapeGML function from the NetPathMiner v1.8.0 package. Similarity to known protein structures was assessed using DALI against a local instance of the PDB database as accessed on May 12, 2022⁸³. Draft OCE structural groups were identified using community detection with the Louvain algorithm implemented in HiDEF⁸⁴ with maximum resolution = 50.0, consensus threshold = 75, persistent threshold = 5, target community number and weight column disabled. Final OCE structural groups were refined manually based on community detection and structural similarity with the PDB database. Groups were named according to the dominant match with PDB or the PDB hit with the highest Z-score, with priority given to proteins of known function. Structural similarity in major OCE families was further supported by TM-scores and RMSD calculated with TM-align⁸⁵ (Supplementary Fig. 12).

Analysis of natural sequence and structure diversity

Structural similarity was assessed using the all-versus-all function in DALI Lite 5.0⁴⁶ and the TM-align program updated on April 12, 2022⁸⁵. Evolutionary distance was based on multiple alignments generated with muscle⁸⁶ and manually curated to keep positions covered at 80% minimum, and assessed using the distmat program in EMBOSS⁸⁷ with Jukes-Cantor correction for multiple substitution. To generate structure similarity trees, Dali Z-scores were converted to distances by calculating the difference to the maximum Z-score in each matrix, and neighbor-joining trees were created with BioNJ. The longest branches were pruned in TreeGraph2⁸⁸. Representative structures were aligned in UCSF Chimera version 1.11.2 build 41376⁸⁹ with MatchMaker for pairwise superposition with the Needleman-Wunsch algorithm and the PAM-250 matrix, secondary structure score (30%) included, secondary structure assignments computed and iteration by pruning long atom pairs until no pair exceeds 0.1 angstrom. To refine the superposition and retrieve the corresponding alignment, the Match->Align tool was used with residue-residue distance cutoff 5.0 angstroms, residue aligned in column if within cutoff of "at least one other", iterate superposition/alignment until convergence, superimpose full columns in stretches of at least 3 consecutive columns. For deep homology analysis, we used BLASTp searches of OCEs (and for Alt-A1 their top five PDB hits) against the NCBI nr database with e-value cutoff 0.05, word size 2, scored using the PAM30 matrix with gap costs 8 for existence and 1 for extension. Amino acid sequences were clustered with a mutual coverage of $\geq 80\%$ and a similarity of $\geq 30\%$ using MMseqs2 version 13.45111⁹⁰. These clusters were then used to create alignments using Decipher version 2.24.0⁹¹. We then used hhsuite3 version 3.3.0⁹² to create HMMs from MMseqs clusters and conduct an all vs all pairwise HMM-HMM comparisons and sequence-HMM comparisons between singletons and clusters. MMseqs clusters were assigned to super-clusters where all members of a super-cluster had a significant HMM similarity to at least one other member at e-value $\leq 1E-5$ (i.e., the connected components of the whole graph).

Mutation scans and determination of residue structural divergence effects

Ancestral OCE sequences were determined with GRASP version 2020.05.05⁹³ using the JTT evolutionary model. Sequence reconstructions at the most ancestral nodes were submitted to AlphaFold2 to obtain the ancestral OCE structures. The residue percent

conservation corresponds to the proportion of modern OCEs harboring the ancestral residue at a given position. A null model for mutation co-occurrence was estimated by randomly shuffling extant polymorphisms 10,000 times and counting the occurrence of all mutation pairs. *P*-values are the product of mutation pair occurrence under the null model exponent the number of observed mutation pairs, followed by Bonferroni correction for multiple testing. Groups of co-selected mutations were identified by community detection applied to the network of co-occurring mutations, using the Louvain algorithm implemented in HiDEF with default parameters. For Alanine scans, amino acids were turned into Alanine one by one. For deletion scans, five consecutive amino acids were deleted in a sliding window approach of step two. For variant shuffling, a repertoire of all variants at a given position detected in the OCE clade was built, these variants were introduced randomly in the ancestral sequence to obtain 100 sequences with an average 5 to 47% positions mutated. For new mutation scans, a repertoire of variant amino acids never found at each position in the OCE clade was built, and mutations were introduced randomly in the ancestral sequence as previously. The structure corresponding to each mutated sequence was predicted using AlphaFold2, and structures were compared using DALIite5 and TMalign. The structure divergence effect SD_i at position *i* was defined as the average $Z_{(\text{ancestor against mutant } i)} - Z_{(\text{ancestor against itself})}$ for alanine and deletion mutants targeting position *i*. The SD compensatory effect at position *i* was defined as $SD_i - [1/\text{number of mutations} * SD_{(\text{shuffle or new mutant including mutation at position } i)}]$ mean-normalized with SD distributions. The net stabilization factor is the difference between SD and SD compensatory effect at position *i*.

Reconstruction of structure and frustration evolutionary paths

Clusters of related OCEs among homologs from NCBI were identified using MMseqs2 (as mentioned previously) and redundant sequences (>97% identity) were filtered out with Cdhit⁹⁴. Signal peptides were removed based on predictions from signalP version 6.0. For these analyses, a selection of the largest alignable groups were retained for six families. Alignments of these clusters were generated with Decipher using default parameters⁹¹. Alignments were manually curated using JalView version 2.11.2.5 to remove sequences that had poorly supported InDels and the N terminus was trimmed to the same location for all cluster members following the signal peptide. Phylogenetic trees were built from these alignments using phyML 3.1 using the approximate likelihood ratio test for branch support, 4 categories of substitution rates and an estimated gamma distribution parameter⁹⁵. Trees were time-calibrated using mid-point rooting and root maximum age calibration with the chronos function of the R package ape v. 5.6-2⁹⁶, with maximum ages inferred using TimeTree 5⁹⁷. Ancestral OCE structures were determined using GRASP and AlphaFold2 as described above. Residue frustration was calculated on AlphaFold rank 1 models using frustratometer version 0.1.0⁵⁸ with long-range electrostatics disabled, mutational calculation mode and sequence separation of 3. Residues were considered highly frustrated for a frustration index (FI) lower than -1. FI variance in OCE clades was calculated as the variance of FIs at a given aligned position for all aligned OCEs. To determine frustration variation along tree branches, the difference between FI in the ancestral protein and FI in the descendant protein was calculated at each aligned position (ΔFI), and ΔFIs summed up along the alignment ($\Sigma \Delta FI$). Structure comparisons were performed using DALIite 5.0 as described above. Structural divergence along branches corresponded to the difference between Dali Z-score for comparison of ancestral structure with itself and ancestral structure with descendant structure. These values were divided by the difference between estimated divergence time of the ancestral and descendant nodes. Trees were rendered using R functions from the package phytools⁹⁸.

Statistics and reproducibility

Statistical analyses were performed with RStudio v1.4.1106 and R x64 4.0.4. Data distributions were modeled using LOESS local regression with confidence or prediction intervals shown. Differences of quantitative data between groups were calculated using 2-sided unpaired *t*-test. Null probability of co-mutation occurrence was determined by shuffling extant variable residues 10,000 times. *P*-values were calculated using Bonferroni correction for multiple testing. For mutational scans, all positions were mutated to alanine or deleted, sets of 100 random sequences with multiple mutations were analyzed to reach a sample size comparable to modern, alanine scan and deletion scan sets. Random multiple mutants were designed by assigning randomly amino-acids independently to each position in proteins. For family-wise analyses, the results from at least two unrelated families spanning >15 species and >65 proteins are reported. Results reported in the manuscript were successfully replicated on all protein families analyzed, with statistics reported in the Source data File.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Source data on zenodo.org under <https://doi.org/10.5281/zenodo.7506581>⁹⁹. Raw data with accession codes and protein identifiers are listed in Source data file 1. Source data are provided with this paper.

References

- Hogenhout, S. A., Van der Hoorn, R. A., Terauchi, R. & Kamoun, S. Emerging concepts in effector biology of plant-associated organisms. *Mol. Plant Microbe Interact.* **22**, 115–122 (2009).
- Hakimi, M. A. & Bougdour, A. Toxoplasma's ways of manipulating the host transcriptome via secreted effectors. *Curr. Opin. Microbiol.* **26**, 24–31 (2015).
- Spassieva, S. D., Markham, J. E. & Hille, J. The plant disease resistance gene Asc-1 prevents disruption of sphingolipid metabolism during AAL-toxin-induced programmed cell death. *Plant J.* **32**, 561–572 (2002).
- Porquier, A. et al. Retrotransposons as pathogenicity factors of the plant pathogenic fungus *Botrytis cinerea*. *Genome Biol.* **22**, 1–19 (2021).
- Kessler, S. C. et al. Victorin, the host-selective cyclic peptide toxin from the oat pathogen *Cochliobolus victoriae*, is ribosomally encoded. *Proc. Natl Acad. Sci. USA* **117**, 24243–24250 (2020).
- He, Q., McLellan, H., Boevink, P. C. & Birch, P. R. J. All roads lead to susceptibility: the many modes of action of fungal and oomycete intracellular effectors. *Plant Commun.* **1**, 100050 (2020).
- Rocafort, M., Fudal, I. & Mesarich, C. H. Apoplastic effector proteins of plant-associated fungi and oomycetes. *Curr. Opin. Plant Biol.* **56**, 9–19 (2020).
- Chen, H., Raffaele, S. & Dong, S. Silent control: microbial plant pathogens evade host immunity without coding sequence changes. *FEMS Microbiol. Rev.* **45**, fuab002 (2021).
- Kamoun, S. Groovy times: filamentous pathogen effectors revealed. *Curr. Opin. Plant Biol.* **10**, 358–365 (2007).
- Kourelis, J. et al. Evolution of a guarded decoy protease and its receptor in solanaceous plants. *Nat. Commun.* **11**, <https://doi.org/10.1038/s41467-020-18069-5> (2020).
- Raffaele, S. & Kamoun, S. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* **10**, 417–430 (2012).
- Möller, M. & Stukenbrock, E. H. Evolution and genome architecture in fungal plant pathogens. *Nat. Rev. Microbiol.* **15**, 756–771 (2017).

13. Fouché, S., Plissonneau, C., Croll, D. & Fouche, S. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Curr. Opin. Microbiol.* **46**, 34–42 (2018).
14. Dong, S., Raffaele, S. & Kamoun, S. The two-speed genomes of filamentous pathogens: Waltz with plants. *Curr. Opin. Genet. Dev.* **35**, 57–65 (2015).
15. Thordal-Christensen, H., Birch, P. R. J., Spanu, P. D. & Panstruga, R. Why did filamentous plant pathogens evolve the potential to secrete hundreds of effectors to enable disease? *Mol. Plant Pathol.* **19**, 781–785 (2018).
16. Kim, K.-T. et al. Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their potential role in host association. *Front. Plant Sci.* **7**, 186 (2016).
17. Haas, B. J. et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**, 393–398 (2009).
18. Bos, J. I. et al. *Phytophthora infestans* effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. *Proc. Natl Acad. Sci. USA* **107**, 9909–9914 (2010).
19. Lanver, D. et al. *Ustilago maydis* effectors and their impact on virulence. *Nat. Rev. Microbiol.* **15**, 409–421 (2017).
20. Navarrete, F. et al. The Pleiades are a cluster of fungal effectors that inhibit host defenses. *PLoS Pathog.* **17**, e1009641 (2021).
21. Brefort, T. et al. Characterization of the largest effector gene cluster of *Ustilago maydis*. *PLoS Pathog.* **10**, e1003866 (2014).
22. Kvitko, B. H. et al. Deletions in the repertoire of *Pseudomonas syringae* pv. tomato DC3000 type III secretion effector genes reveal functional overlap among effectors. *PLoS Pathog.* **5**, e1000388 (2009).
23. Leisen, T. et al. Multiple knockout mutants reveal a high redundancy of phytotoxic compounds contributing to necrotrophic pathogenesis of *Botrytis cinerea*. *PLoS Pathog.* **18**, 1–28 (2022).
24. John, E. et al. Variability in an effector gene promoter of a necrotrophic fungal pathogen dictates epistasis and effector-triggered susceptibility in wheat. *PLoS Pathog.* **18**, e1010149 (2022).
25. Martel, A. et al. Metaeffector interactions modulate the type III effector-triggered immunity load of *Pseudomonas syringae*. *PLoS Pathog.* **18**, e1010541 (2022).
26. Lazar, N. et al. A new family of structurally conserved fungal effectors displays epistatic interactions with plant resistance proteins. *PLoS Pathog.* **18**, e1010664 (2022).
27. Snelders, N. C. et al. Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins. *Nat. Plants* **6**, 1365–1374 (2020).
28. Snelders, N. C., Petti, G. C., van den Berg, G. C. M., Seidl, M. F. & Thomma, B. P. H. J. An ancient antimicrobial protein co-opted by a fungal plant pathogen for in planta mycobiome manipulation. *Proc. Natl Acad. Sci. USA* **118**, e2110968118 (2021).
29. Franceschetti, M. et al. Effectors of filamentous plant pathogens: commonalities amid diversity. *Microbiol. Mol. Biol. Rev.* **81**, e00066–16 (2017).
30. Macquet, J., Mounichetty, S. & Raffaele, S. Genetic co-option into plant–filamentous pathogen interactions. *Trends Plant Sci.* **27**, 1144–1158 (2022).
31. de Guillen, K. et al. Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi. *PLoS Pathog.* **11**, 1–27 (2015).
32. Praz, C. R. et al. *AvrPm2* encodes an RNase-like avirulence effector which is conserved in the two different specialized forms of wheat and rye powdery mildew fungus. *N. Phytol.* **213**, 1301–1314 (2017).
33. Di, X. et al. Structure–function analysis of the *Fusarium oxysporum* *avr2* effector allows uncoupling of its immune-suppressing activity from recognition. *N. Phytol.* **216**, 897–914 (2017).
34. Yu, D. S. et al. The structural repertoire of *Fusarium oxysporum* f. sp. *lycopersici* effectors revealed by experimental and computational studies. *Elife* **12**, RP89280 (2023).
35. Guyon, K., Balagué, C., Roby, D. & Raffaele, S. Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen *Sclerotinia sclerotiorum*. *BMC Genom.* **15**, 336 (2014).
36. Seong, K. & Krasileva, K. V. Computational structural genomics unravels common folds and novel families in the secretome of fungal phytopathogen *Magnaporthe oryzae*. *Mol. Plant-Microbe Interact.* **34**, 1267–1280 (2021).
37. Rocafort, M. et al. The *Venturia inaequalis* effector repertoire is dominated by expanded families with predicted structural similarity, but unrelated sequence, to avirulence proteins from other plant-pathogenic fungi. *BMC Biol.* **20**, 246 (2022).
38. Seong, K. & Krasileva, K. V. Prediction of effector protein structures from fungal phytopathogens enables evolutionary analyses. *Nat. Microbiol.* **8**, 174–187 (2023).
39. Jones, D. A. B., Moolhuijzen, P. M. & Hane, J. K. Remote homology clustering identifies lowly conserved families of effector proteins in plant-pathogenic fungi. *Microb. Genomics* **7**, 000637 (2021).
40. Białas, A. et al. Lessons in effector and NLR biology of plant-microbe systems. *Mol. Plant-Microbe Interact.* **31**, 34–45 (2018).
41. Oikawa, K. et al. The blast pathogen effector AVR-Pik binds and stabilizes rice heavy metal-associated (HMA) proteins to co-opt their function in immunity. *bioRxiv* **12**, 406389 (2020).
42. Park, C.-H. et al. The *Magnaporthe oryzae* effector *AvrPiz-t* targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern-triggered immunity in rice. *Plant Cell* **24**, 4748–4762 (2012).
43. Corsi, B. et al. Genetic analysis of wheat sensitivity to the ToxB fungal effector from *Pyrenophora tritici-repentis*, the causal agent of tan spot. *Theor. Appl. Genet.* **133**, 935–950 (2020).
44. Bloom, J. D. et al. Evolution favors protein mutational robustness in sufficiently large populations. *BMC Biol.* **5**, 29 (2007).
45. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
46. Holm, L. & Park, J. DALI: a workbench for protein structure comparison. *Bioinformatics* **16**, 566–567 (2000).
47. Antuch, W., Guntert, P. & Wuthrich, K. Ancestral β -crystallin precursor structure in a yeast killer toxin. *Nat. Struct. Biol.* **3**, 662–665 (1996).
48. Li, N. et al. Structure of *Ustilago maydis* killer toxin KP6 α -subunit. A multimeric assembly with a central pore. *J. Biol. Chem.* **274**, 20425–20431 (1999).
49. Pennington, H. G. et al. The fungal ribonuclease-like effector protein CSEP0064/BEC1054 represses plant immunity and interferes with degradation of host ribosomal RNA. *PLoS Pathog.* **15**, e1007620 (2019).
50. Gupta, G. D., Bansal, R., Mistry, H., Pandey, B. & Mukherjee, P. K. Structure-function analysis reveals *Trichoderma virens* Tsp1 to be a novel fungal effector protein modulating plant defence. *Int. J. Biol. Macromol.* **191**, 267–276 (2021).
51. Liu, W. et al. Mutational analysis of the *Verticillium dahliae* protein elicitor PevD1 identifies distinctive regions responsible for hypersensitive response and systemic acquired resistance in tobacco. *Microbiol. Res.* **169**, 476–482 (2014).
52. Chruszcz, M. et al. *Alternaria alternata* allergen Alt a 1: a unique β -barrel protein dimer found exclusively in fungi. *J. Allergy Clin. Immunol.* **130**, 241 (2012).
53. Liu, M. et al. Crystal structure analysis and the identification of distinctive functional regions of the protein elicitor Mohrip2. *Front. Plant Sci.* **7**, 1103 (2016).
54. Outram, M. A. et al. The crystal structure of SnTox3 from the necrotrophic fungus *Parastagonospora nodorum* reveals a unique

- effector fold and provides insight into Snn3 recognition and pro-domain protease processing of fungal effectors. *N. Phytol.* **231**, 2282–2296 (2021).
55. Wang, X., Minasov, G. & Shoichet, B. K. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* **320**, 85–95 (2002).
56. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA* **103**, 5869–5874 (2006).
57. Ferreiro, D. U., Hegler, J. A., Komives, E. A. & Wolynes, P. G. Localizing frustration in native proteins and protein assemblies. *Proc. Natl Acad. Sci. USA* **104**, 19819–19824 (2007).
58. Rausch, A. O. et al. Frustratometer: an R-package to compute local frustration in protein structures, point mutants and MD simulations. *Bioinformatics* **37**, 3038–3040 (2021).
59. de Guillen, K. et al. Structural genomics applied to the rust fungus *Melampsora larici-populina* reveals two candidate effector proteins adopting cystine knot and NTF2-like protein folds. *Sci. Rep.* **9**, 1–12 (2019).
60. Schmitt, M. J. & Breinig, F. The viral killer system in yeast: from molecular biology to application. *FEMS Microbiol. Rev.* **26**, 257–276 (2002).
61. Koltin, Y. & Day, P. R. Inheritance of killer phenotypes and double stranded RNA in *Ustilago maydis*. *Proc. Natl Acad. Sci. USA* **73**, 594–598 (1976).
62. Guan, Y. et al. Horizontally acquired fungal killer protein genes affect cell development in mosses. *Plant J.* **113**, 665–676 (2022).
63. Czislowski, E. et al. Investigation of the diversity of effector genes in the banana pathogen, *Fusarium oxysporum* f. sp. *cubense*, reveals evidence of horizontal gene transfer. *Mol. Plant Pathol.* **19**, 1155–1171 (2018).
64. McDonald, M. C. et al. Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA between three fungal wheat pathogens. *MBio* **10**, e01515–e01519 (2019).
65. Zhang, Q. et al. Horizontal gene transfer allowed the emergence of broad host range entomopathogens. *Proc. Natl Acad. Sci. USA* **116**, 7982–7989 (2019).
66. Lenarčič, T. et al. Eudicot plant-specific sphingolipids determine host selectivity of microbial NLP cytolysins. *Science* **358**, 1431–1434 (2017).
67. Seidl, M. F. & Van Den Ackerveken, G. Activity and phylogenetics of the broadly occurring family of microbial nep1-like proteins. *Annu. Rev. Phytopathol.* **57**, 367–386 (2019).
68. Snelders, N. C., Rovenich, H. & Thomma, B. P. H. J. Microbiota manipulation through the secretion of effector proteins is fundamental to the wealth of lifestyles in the fungal kingdom. *FEMS Microbiol. Rev.* **46**, fuac022 (2022).
69. Oome, S. et al. Nep1-like proteins from three kingdoms of life act as a microbe-associated molecular pattern in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **111**, 16955–16960 (2014).
70. Denton-Giles, M. et al. Conservation and expansion of a necrosis-inducing small secreted protein family from host-variable phytopathogens of the Sclerotiniaceae. *Mol. Plant Pathol.* **21**, 512–526 (2020).
71. Jemth, P., Mu, X., Engström, Å. & Dogan, J. A frustrated binding interface for intrinsically disordered proteins. *J. Biol. Chem.* **289**, 5528–5533 (2014).
72. Freiburger, M. I., Wolynes, P. G., Ferreiro, D. U. & Fuxreiter, M. Frustration in fuzzy protein complexes leads to interaction versatility. *J. Phys. Chem. B* **125**, 2513–2520 (2021).
73. Stelzl, L. S. et al. Local frustration determines loop opening during the catalytic cycle of an oxidoreductase. *Elife* **9**, 1–27 (2020).
74. Jeblick, T. et al. Botrytis hypersensitive response inducing protein 1 triggers noncanonical PTI to induce plant cell death. *Plant Physiol.* **191**, 125–141 (2023).
75. Naganathan, A. N. & Kannan, A. A hierarchy of coupling free energies underlie the thermodynamic and functional architecture of protein structures. *Curr. Res. Struct. Biol.* **3**, 257–267 (2021).
76. Tzul, F. O., Vasilchuk, D. & Makhatadze, G. I. Evidence for the principle of minimal frustration in the evolution of protein folding landscapes. *Proc. Natl Acad. Sci. USA* **114**, E1627–E16322 (2017).
77. Nobrega, R. P. et al. Modulation of frustration in folding by sequence permutation. *Proc. Natl Acad. Sci. USA* **111**, 10562–10567 (2014).
78. Teufel, M., Zajc, C. U. & Traxlmayr, M. W. Engineering strategies to overcome the stability-function trade-off in proteins. *ACS Synth. Biol.* **11**, 1030–1039 (2022).
79. Nielsen, H. in *Methods in Molecular Biology* Vol. 1611, 59–73 (Humana Press Inc., 2017).
80. Sperschneider, J. & Dodds, P. N. EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol. Plant-Microbe Interact.* **35**, 146–156 (2022).
81. Walsh, I., Martin, A. J. M., Di Domenico, T. & Tosatto, S. C. E. ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* **28**, 503–509 (2012).
82. Holm, L. DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140 (2020).
83. Berman, H. M. et al. The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
84. Zheng, F. et al. HiDeF: identifying persistent structures in multiscale ‘omics data. *Genome Biol.* **22**, 1–15 (2021).
85. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
86. Madeira, F. et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).
87. Rice, P., Longden, L., Bleasby, A., Longden, I. & Bleasby, A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).
88. Stöver, B. C. & Müller, K. F. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinform.* **11**, 7 (2010).
89. Pettersen, E. F. et al. UCSF Chimera a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
90. Hauser, M., Steinegger, M. & Söding, J. MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics* **32**, 1323–1330 (2016).
91. Wright, E. S. DECIPHER: Harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinform.* **16**, 322 (2015).
92. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinform.* **20**, 473 (2019).
93. Foley, G. et al. Engineering indel and substitution variants of diverse and ancient enzymes using Graphical Representation of Ancestral Sequence Predictions (GRASP). *PLOS Comput. Biol.* **18**, e1010633 (2022).
94. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
95. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
96. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
97. Kumar, S. et al. TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* **39**, msac174 (2022).
98. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

99. Derbyshire, M. & Raffaele, S. Supplementary material for 'Surface frustration re-patterning underlies the structural landscape and evolvability of fungal orphan candidate effectors'. <https://doi.org/10.5281/ZENODO.7506581> (2023).

Acknowledgements

SR was supported by the French Laboratory of Excellence project 'TULIP' (ANR-10-LABX-41; ANR-11-IDEX-0002-02), l'Agence Nationale pour la Recherche (ANR-19-CE20-15, ANR-21-CE20-10, ANR-21-CE20-30), and the INRAE. S.R. is grateful the LIPME bioinformatics team and especially Ludovic Legrand for providing help and computing and storage resources. MCD was supported by a co-investment between the Grains Research and Development Corporation of Australia and Curtin University on grant CUR00023. MCD would like to acknowledge that this work was supported by resourced provided by the Pawsey Super-computing Research Centre with funding from the Australian Government and the Government of Western Australia.

Author contributions

S.R. conceptualized the study. M.C.D. and S.R. designed, performed and interpreted analyses, wrote and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-40949-9>.

Correspondence and requests for materials should be addressed to Sylvain Raffaele.

Peer review information *Nature Communications* thanks the anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023