



**HAL**  
open science

## Exploring a diverse world of effector domains and amyloid signaling motifs in fungal NLR proteins

Jakub W Wojciechowski, Emirhan Tekoglu, Marlena Gąsior-Glogowska, Virginie Coustou, Natalia Szulc, Monika Szefczyk, Marta Kopaczyńska, Sven J Saupe, Witold Dyrka

### ► To cite this version:

Jakub W Wojciechowski, Emirhan Tekoglu, Marlena Gąsior-Glogowska, Virginie Coustou, Natalia Szulc, et al.. Exploring a diverse world of effector domains and amyloid signaling motifs in fungal NLR proteins. 2022. hal-03854981

**HAL Id: hal-03854981**

**<https://hal.science/hal-03854981>**

Preprint submitted on 16 Nov 2022

**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.

# Exploring a diverse world of effector domains and amyloid signaling motifs in fungal NLR proteins

Jakub W. Wojciechowski\*<sup>1</sup>, Emirhan Tekoglu\*<sup>†2, 3</sup>, Marlena Gašior-Głogowska<sup>1</sup>,  
Virginie Coustou<sup>4</sup>, Natalia Szulc<sup>1</sup>, Monika Szeferczyk<sup>5</sup>, Marta Kopaczyńska<sup>1</sup>,  
Sven J. Saupe<sup>‡4</sup>, and Witold Dyrka<sup>‡1</sup>

<sup>1</sup>Politechnika Wroclawska, Wydział Podstawowych Problemów Techniki,  
Katedra Inżynierii Biomedycznej, Poland

<sup>2</sup>Yıldız Teknik Üniversitesi, Biyomühendislik Bölümü, İstanbul, Turkey

<sup>3</sup>Politechnika Wroclawska, Wydział Chemiczny, Poland

<sup>4</sup>Institut de Biochimie et de Génétique Cellulaire, UMR 5095 CNRS, Université  
de Bordeaux, 33077 Bordeaux CEDEX, France

<sup>5</sup>Politechnika Wroclawska, Wydział Chemiczny, Katedra Chemii Bioorganicznej,  
Poland

## Abstract

NLR proteins are intracellular receptors constituting a conserved component of the innate immune system of multicellular organisms. In fungi, NLRs are characterized by high diversity of architectures and presence of amyloid signaling. Here, we explore the diverse world of effector and signaling domains of fungal NLRs using state-of-the-art bioinformatic methods including MMseqs2 for fast clustering, probabilistic context-free grammars for sequence analysis, and AlphaFold2 deep neural networks for structure prediction. In addition to substantially improving the overall annotation, especially in basidiomycetes, the study identifies novel domains and reveals the structural similarity of MLKL-related HeLo- and Goodbye-like domains forming the most abundant superfamily of fungal NLR effectors. Moreover, compared to previous studies, we found several times more amyloid motifs, including novel families, and validated aggregating and prion-forming properties of the most abundant of them *in vitro* and *in vivo*. Also, through an extensive *in silico* search, the NLR-associated amyloid signaling is for the first time identified in basidiomycetes. The emerging picture highlights similarities and differences in the NLR architectures and amyloid signaling in ascomycetes, basidiomycetes and other branches of life.

**Keywords:** *NLR proteins, amyloid signaling motif, sequence motifs, remote homology, structure prediction, non-self recognition, fungi*

---

\*Contributed equally

<sup>†</sup>Present address: Koc University, School of Medicine, İstanbul, Turkey

<sup>‡</sup>Corresponding author

## Background

**NLR proteins.** All living organisms possess an immune system allowing them to cope with viral or cellular pathogens. Among the central and conserved components of the innate immune system in animals and plants are the NLR proteins. NLRs are intracellular immune receptors that induces various host responses including regulated cell death upon the detection of non-self cues [1, 2, 3]. A typical NLR protein functions following a ligand-induced oligomerisation and activation process. Its tripartite domain architecture displays 1) a central Nucleotide-binding and Oligomerization Domain (NOD), 2) a C-terminal domain composed of superstructure forming repeats that is typically involved in detection of non-self cues in the form of DAMPs or MAMPs (Damage- or Microbe-Associated Molecular Patterns) and 3) a N-terminal effector domain whose activation induces various downstream host responses including regulation of the infected cell death [4, 5, 6, 7, 8]. While historically, NLRs were mostly studied within the animal and plant kingdoms (as Nod-Like Receptors and NBS-LRR Receptors respectively) [9, 10], their homologs were identified in bacteria and fungi [4].

In fungi, homologs of NLR proteins were initially identified in the context of the study of a non-self recognition process termed heterokaryon incompatibility (HI) [11]. This reaction occurs in filamentous fungi in the event of the fusion (anastomosis) of the hyphae of genetically incompatible individuals, resulting in the death of mixed fusion cells [12, 13]. Incompatibility prevents in particular the transmission of mycoviruses between isolates during the anastomosis events. In *Podospora anserina*, HET-E, one of the proteins controlling heterokaryon incompatibility is a homolog of NLR proteins (although its N- and C-terminal domains differ from those known in animals and plants, a situation typical for NLR architecture proteins outside of the plant and animal kingdom [4, 14, 15]). Its central NOD domain is one of the original founding members used to define the NACHT domain (Pfam PF05729) common in animal NLRs (the H in the NACHT acronym stands for HET-E) [10, 16]. The C-terminal domain of HET-E protein, built of hypervariable WD40 repeats recognizes a non-self cue, here polymorphic variants of a host protein termed HET-C, a glycolipid transfer protein universally conserved in eukaryotes that could represent a pathogen effector target [17]. In such event, the N-terminal HET domain is activated which eventually leads to regulated cell death [17]. The HET domain (PF06985) is a cell death inducing domain with a remote homology to TIR domains [14]. Several other fungal cell death inducing incompatibility pathways in *Podospora* and other species are controlled by NLR proteins [18, 5]. Yet, apparently only a small fraction of the existing fungal NLRs appear involved in heterokaryon incompatibility and it is proposed that these proteins have more general function in immune defense and establishment of symbiotic interactions in fungi [19, 5]. Indeed, NLR proteins are abundant in multicellular filamentous fungi (no NLR protein was found in unicellular yeasts). In a recent study, a total of about 36 000 NLR proteins have been found in around 880 strains of over 560 species of fungi with on average 57 NLRs per genome and numerous species displaying hundreds of NLR genes [14, 5].

In terms of domain annotation fungal NLRs differ from their typical animal and plant counterparts. Unlike more homogenous NLR proteins in animals and plants, the central domain of fungal NLRs can be either of the NACHT [10] or the NB-ARC type (PF00931) [9]. Then fungal NLRs display ankyrin repeats (ANK, Pfam CL0465), tetratricopeptide repeats (TPR, CL0020) and beta-propellers of the WD40 meta-family (CL0186) in place of the LRR repeats found in most animal and plant NLRs. The NBS-TPR architecture was proposed to correspond to the ancestral architecture whilst NLR proteins in multicellular bacteria also typically display TPR, ANK or WD

repeats [4, 14, 15, 20]. Consistent with a role in immune defense C-terminal repeated domains of fungal NLRs display marks of positive selection and are highly variable [21, 14, 18]. In addition, the C-terminal domains show original modes of functional diversification. First, about 1/6 of these C-terminal repeat domains consist of highly similar repeats with only a few highly variable positions under positive selection [22, 14]. These repeats arrays with high internal similarity are hypervariable loci in which individual repeats are exchanged and reshuffled resulting in functional diversification [21, 22]. High internal similarity of repeats is both a cause and a result of an unequal crossing over mechanism, a process which is 5-6 orders of magnitude faster than the point mutation [23]. Then, in the truffle *Tuber melanosporum* a superfamily of NACHT-ANK NLR encoding genes displays dozens of 3 bp minixons whose alternative splicing can considerably diversify the repertoire of potential C-terminal recognition domain [24]. These striking modes of recognition domain diversification are consistent with the proposed role of NLR proteins in the immune response, as capability of quickly adapting to evolving pathogens is a condition of success in the constant arms race against them [21].

For about 50% of fungal NLR proteins, N-terminal domain annotations could be determined with the Pfam [25] and similar HMM profiles [14], which make up for 12-13 major meta-families [14, 5]. Functionally, the characterized N-terminal domains belong to three basic types: enzymatic, signaling, and regulated cell death induction [26]. Out of 72 possible architectures made with the most common domain families (12 types of N-terminal domains, 2 types of central domains and 3 clans of C-terminal domains), as many as 32 were identified in fungal proteomes [14]. Interestingly, in about 20 cases, the closest orthologs of the central domain sequences were bound to different N-terminal domains (including in two different strains of the same species). Moreover, the maximum-likelihood phylogenetic trees generated separately for the N-terminal and central domains were mutually incompatible, and distribution of the N-terminal domains over the branches of central domains trees generated for selected species was scattered. Together with a relatively high number of NLRs without ortholog in other strains of the same species, these findings indicate high plasticity of the architecture of NLR proteins and the occurrence of the *death-and-birth evolution* process [14, 5].

**Amyloid signaling motifs.** Another notable feature of fungal NLRs is the occurrence of amyloid-forming motifs at their N-termini [26]. A series of studies derived from the characterization of *Podospora anserina* [Het-s] prion protein, which controls regulated cell death in the context of heterokaryon incompatibility, has revealed that a fraction of the fungal NLR employ amyloid signaling to activate downstream cell death effector domains [27, 26]. The paradigmatic example of such amyloid NLR signalosomes is the HET-S/NWD2 two-component system of *P. anserina*. HET-S encodes a cell death execution protein with a globular N-terminal HeLo domain (PF14479) and a C-terminal amyloid forming prion domain composed of two elementary repeats r1 and r2 which are able adopt a specific  $\beta$ -solenoid amyloid fold [28, 29, 30, 31]. Amyloid transconformation of the C-terminal domain induces activation of the HeLo domain which turns into a pore-forming toxin. NWD2 is a NLR, encoded by the gene immediately adjacent to *het-S*, and displays at its N-terminus a motif termed r0 which is homologous to the elementary r1 and r2 repeats [32, 27]. When activated by their cognate ligand, engineered variants of NWD2 are capable of triggering transconformation of HET-S and to induce its toxicity. In this system, activation of the NLR leads to amyloid folding of its N-terminus which then serves as template to activate a cognate cell death execution protein [26].

The r0, r1, and r2 motifs, collectively referred to as the HET-s *motif*, represent one of the best studied examples of an amyloid signaling motif (ASM). Homologs of the HET-s motif can be grouped in 5 subclasses (collectively denoted as HET-s Related Amyloid Motifs or HRAM) [33], which co-occur in N-termini of fungal NLR proteins and in C-termini of HeLo [34, 35, 29] and HeLo-like (PF17111) proteins [32, 14] encoded by genes adjacent to NLR-encoding genes in the genome. In some organisms, two or three subclasses of HRAMs exist simultaneously, which allows for maintaining distinct signaling pathways [33, 36].

There are two other families of Fungal Amyloid Signaling Sequences (FASS) with similar functionality in the NLR protein system, namely sigma (named after the  $\sigma$  prion, which contains this motif [37]) and PP (pseudopalindromic due to the amino acid pattern NxGxQxGxN at its core) [32]. The PP motif bears significant resemblance to the mammalian RHIM motif [38, 39, 40] with remote homologs also in multicellular bacteria [20]. In addition, a recent *in silico* analysis of over 100,000 available bacterial genomes in search of sequence motifs repeated in adjacent genes encoding the Bell (bacterial homolog of fungal HeLo) and NLR proteins revealed ten families of Bacterial Amyloid Signal Sequences (BASS) widespread in multicellular Actinomycetes, Cyanobacteria and in Archaea [20]. Despite their sequence-level diversity, at least some if not all known BASS and FASS motifs are believed to share the beta-arch fold [41, 42, 43].

When compared to the NLR proteins in plant and animal kingdoms, the fungal NLR proteins display larger diversity of architectures. In addition, NLR-associated amyloid signaling appears specific to fungal and bacterial kingdoms although amyloid-motifs also occur in immune pathways in animals [44, 45]. The dominant view, until recently, was that the architecture and immunological function of NLR proteins in plants and animals resulted from the convergent evolution [15]. However, higher diversity of NLRs in fungi than in animals and plants, as well as presence of NLRs in multicellular prokaryotes [4, 20] suggest the early evolutionary origins of the architecture and the immune function of NLR proteins [5, 26]. Exploration of the diversity of fungal NLRs is an important asset for deciphering of the potential roles of these immune receptors in fungal biology in addition to their documented role in cell death related to incompatibility. In addition, comparative studies of NLRs in the different kingdoms can provide a more global view of the long term evolution of these central components of immunity in both microbes and macro-organisms. The aim of the current study is to improve the annotation and characterization of the vast ensemble of N-terminal domain of fungal NLRs with special focus on short domains (shorter than 150 amino acids) and amyloid-like motifs.

## Results

### Overview of N-terminal domains of fungal NLRs

In a previous study, we identified 36 141 NLR proteins in 487 fungal strains out of a total of 882 strains of 561 species for which the genome sequencing data was then available [20]. The ratio between the number of sequences from Ascomycota and Basidiomycota was 3.12:1 (27 152 to 8708). The N-terminus, as delimited according to NACHT or NB-ARC query match, was at least 20 amino-acids long in 32 962 proteins. Among these N-termini, 18 674 (57%) were annotated using either the Pfam [25] or our inhouse profiles [14]. In order to improve the annotation coverage,

we clustered the set of N-termini with MMseqs2 [46] and then, for each cluster with more than 20 members (21 758 N-termini in 127 clusters, 15 105 already annotated), searched for homologs in UniRef30 [47, 48] and subsequently in Pfam using HHblits [49]. The clustering required mutual coverage of at least 80% of sequence length, and annotations were only assigned to sequences which overlapped at least 50% of the match to the Pfam profile. Moreover, only the matches including at least 50% of the Pfam profile length were retained to avoid excessive number of false positive hits. The procedure led to assigning the Pfam-based annotations to 3003 additional N-termini. This increased the annotation coverage from 69% to 82% (17 902 sequences) in the 20+ clusters, and from 57% to 66% (21 677 sequences) in the entire N-termini dataset.

The corpus of fungal NLR N-termini can be broadly divided into four main super-architectures (Fig. 1a). The first of them consists of the shortest N-termini, up to roughly 50 amino acid in length, which mainly comprise a direct N-terminal extension of the nucleotide binding domain. The second architecture adds the amyloid signaling motifs (ASM), typically made of 20–30 residues, which makes the entire N-terminus 70–120 amino-acids long. The third and the most frequent architecture consists of a single effector domain, while the fourth comprises with multiple domains. The length distribution of N-terminal sequences varies significantly with regard to the fungal phylum (Fig. 1b): while Basidiomycota are over-represented among short N-termini (below 100 amino acids), Ascomycota make up for 85% of domains longer than 200 amino acids.

The Pfam annotation coverage is not evenly distributed. While almost 90% of longer N-terminal domains (200 amino acids or more) are at least partially annotated, the figure is below 40% for the 100–200 amino acids range, and — not surprisingly — a few percent for domains shorter than 100 amino acids, which constitute 1/4 of all NLR N-termini (Fig. 1c). Also, sequences from Ascomycota are more completely annotated (72%) than Basidiomycota (23%) even though our the clustering-based annotation scheme increased coverage of the latter branch roughly twice (Fig. 1d). This inequality holds as well when N-termini in the same length ranges (above 100aa) are compared in both branches. When considering only 127 clusters from MMseqs2, Pfam annotations were found in more than 80% sequences with UniRef100 homologs outside the Fungi kingdom, but only in about 20% sequences with fungal-only homologs (Fig. 1e).

**Major differences.** Pfam annotations for fungal NLR N-termini are listed in Tab. 1 and summarized in Fig. 1f, and in Fig. S1 in Supplementary File 1. Vast majority of newly added annotations belonged to domain families already described as fungal NLR effectors (Tab. 1). In addition to domains reported in our previous surveys [14, 20], this involved also the Crinkler domain of the Ubiquitin clan only recently included in Pfam [50, 51, 52, 53, 54]. In our dataset the domain was found in two mycorrhizal species from two different phyla: *Serendipitia vermifera* (Basidiomycota) and *Rhizophagus irregularis* (Mucormycota). The clustering-based search revealed also one new annotation (assigned to 24 sequences), the Sterile Alpha Motif family SAM\_Ste50p [55]. SAM motifs are involved in homologous and heterologous protein-protein interactions [56].

On the other hand, no cluster matched the previously reported RelA\_SpoT [57], PKinase [58], CHAT [59] and TIR [60, 61] profiles with sufficiently high hit coverage. TIR is a remote homologue of HET [14]. Interestingly, in the profile-sequence search PKinase was found to be N-terminally attached to another N-terminal domain in 29 cases, therefore it is not unlikely that the remaining single PKinase hits are also partial annotations. While no cluster was annotated as C2 only [62], a huge number of sequences fell into the C2 Goodbye-like class [32, 14] (356 hits), the architecture which is specific to Agaricomycetes. The Goodbye-like domain was found also in other double

domain architectures of NLR N-termini (Tab. 1).

**Major families.** While the clustering-based approach found less SesB-like annotations than the profile-sequence search, this is likely due to the fact that the former procedure used only Pfam AB\_Hydrolase clan [63] entries, while the latter method also relied on a sensitive inhouse SesB-like profile [14, 5]. On the contrary, the new approach increased the number of the Purine and Uridine Phosphorylase (PUP) superfamily annotations, mostly due to the matches to the purine Nucleoside Permease (NUP) profile [64]. In addition, 45 sequences from 27 various Pezizomycotina species comprised of about 50 amino-acid long C-terminus of PUP. The fragment forms an intrinsic part of the entire PNP\_UDP\_1 fold (cf. pdb:6po4B, residues 176–234). Both domain families are common in ascomycetal NLRs (13–14% each) but are virtually (SesB-like) or completely (PUP) missing from basidiomycotal NLRs (Fig. S1). The same is true of the HET domain present as N-terminal domain in ascomycetes but not in basidiomycetes.

Notably, we found clusters with apparently overlapping HeLo/HeLo-like and HeLo-like/Goodbye-like domain annotations. The latter situation was found in Basidiomycota and mostly involved sequences annotated as MLKL\_NTD according to Conserved Domain Database (CDD) [65]. Human MLKL is an executioner domain homologous to fungal HeLo and bacterial Bell domains [32, 40, 20]. Moreover, there were additional basidiomycotal clusters with CDD MLKL\_NTD annotation and/or with Pfam HeLo- or Goodbye-like annotations just below the assignment threshold, surmounting to a total of 600 basidiomycotal MLKL-like (BaMLKL) sequences. This makes the superfamily of Goodbye/HeLo/MLKL\_NTD-like domains the most frequent in Basidiomycota (nearly 2000 sequences, around 1/4 of all), similarly to Ascomycota (Fig. S1).

Overall, the three most abundant domain classes, foremost the Goodbye/HeLo/MLKL-like followed by the SesB-like and PUP families, account for 60% of fungal NLR N-termini (Fig. 1f).

## Relation between HeLo-, Goodbye- and basidiomycotal MLKL-likes

To gain more insight in fungal MLKL-likes, we analyzed the largest cluster (OBZ65626, 106 sequences) with the overlapping Goodbye-like and HeLo-like annotations assigned through the HHblits-based procedure. Several sequences in the cluster received also hits from various MLKL-related Pfam profiles when sequences were searched individually (sequence and domain E-values of  $1e - 3$ , Fig. 2a). Not surprisingly, the multiple sequence alignment of the cluster closely matched (HHpred [66, 67] probability above 98%) the sequence of human MLKL executioner domain with an experimentally solved spatial structure (pdb:6vzo [68], Fig. 2b). In fact the MLKL domain is virtually perfectly aligned with the Helo\_like\_N profile match, while the related SesA profile match is slightly shorter. At the same time, the matches to the two Goodbye-like profiles, Goodbye and NACHT\_N [14], are both shifted N-terminally with regard to the MLKL-like domain resulting in a partial overlap, significantly longer for NACHT\_N. Importantly, the multiple sequence alignment is well conserved for the combined stretch of Goodbye- and HeLo-like matches regardless of Pfam annotations of individual sequences (Fig. 2a).

Then, we attempted structure prediction for the largest MLKL-like clusters using AlphaFold2 [69] through the ColabFold advanced notebook [70]. The predictions were carried out solely using multiple sequences alignments of each cluster. Except for the largest HeLo-like cluster, all other predictions resulted in very good quality models (IDDT around 0.80) sharing a four-helix core (Fig. 2c), which is characteristic to the solved MLKL structure. When aligned to the latter using

TMalign [71], the predicted models achieved TM-scores between 0.51 to 0.64. The most notable difference between structural models obtained for various clusters is an additional N-terminal helix in basidiomycotal MLKL\_NTD homologs and Goodbye-likes (hG in Fig. 2), not found in MLKL and HeLo-likes. However, Goodbye-like models present longer and more complex N-terminal extension than BaMLKLs. Also, Goodbye-likes lack a short perpendicular helix (hM) between helices h3 and h4, which seems to be a common feature of human and basidiomycotal MLKLs and HeLo-likes (Fig. 2c).

Taken together, these analyses indicate that although Goodbye-like profiles share a core region with the MLKL bundle and HeLo and Helo-like profiles, they also differ by the presence of an N-terminal extension ahead of the region corresponding to the first helix in MLKL/RPW8/HeLo proteins. Considering the critical role of this region in the membrane targeting activity of these animal, plant and fungal proteins, further experimental investigation are needed before a potential cell death inducing activity can be firmly attributed to Goodbye-like profiles [31, 72, 73].

## Unannotated longer N-termini

Roughly half of the 127 clusters with at least 20 member sequences did not get any Pfam annotation through the HHblits procedure. We carefully examined clusters with at least 10 non-redundant sequences (identity threshold of 70% or nr70) and median length above 100 amino acids. Apart from the MLKL-related clusters described above, there were five such clusters with a total of 195 sequences (127 at nr70) corresponding to four candidate domain families (Tab. 2). All five clusters were associated with NACHT NLRs.

Two clusters (with a total of 94 sequences) consisted of homologous (HHalign [49] E-value of  $6e - 44$ ) relatively long domains (N-terminal length above 500 amino acids) from Pezizomycotina (e.g. KEY84097 protein from *Aspergillus fumigatus*) and Mortierellomycetes (e.g. KFH66451 from *Podila verticillata*). Structure prediction with AlphaFold2 [69, 70] revealed the domain is made of multiple alpha-helices forming two stretches of the alpha solenoid-like structure (Fig. S2ab). NLRs with this N-terminal domain usually display C-terminal repeats of type HEAT (from the TPR clan) in Ascomycota, and WD40 in Mucormycota. Homologous proteins were also found (through the web-based profile HMM search) in bacteria, mainly in *Mycoavidus cysteinexigens*. As this betaproteobacteria is an endosymbiont of *Linnemannia (Mortierella) elongata* AG-77, a fungus with the largest number of these proteins [2], this may suggest possibility of the horizontal gene transfer.

The second most abundant unannotated cluster included 41 sequences (32 non-redundant at nr70) from *Agaricus bisporus* and *Leucoagaricus sp. SymC.cos* (representative protein XP\_007333708 from *A. bisporus*), yet the profile HMM search revealed also single homologs in *Coprinopsis marcescibilis*, yeast *Saprochaete ingens* and protozoan parasite *Eimeria burnetti*. Interestingly, the latter species is a host of *Totiviridae* RNA virus similar to fungal viruses [74] reported also in Agaricaceae [75, 76]. These NLR N-termini (median length of 246 amino acids) consist of proline-rich disordered region and stretches of amyloid-like composition, including a conserved motif in C-terminus. The yeast homolog (non-NLR) is partially annotated as the glycogen recognition site of AMP-activated protein kinase AMPK1\_CBM (PF16561) [77]. Due to the central disordered part, poorly alignable, no reliable structure prediction was possible.

Third most populated unannotated cluster consisted of 31 sequences (15 at nr70, representative protein KIL58680) with the median length of 210 amino acids, which are specific to the *Amanita*

*muscaria* strain Koide BX008. In NLRs the domain is typically associated with C-terminal WD repeats. Again, no reliable structure was predicted.

The final candidate effector domain identified in this study was found in 29 sequences (19 at nr70) from various Pezizomycotina species (representative protein PQE30996 from *Rutstroemia* sp. NJR-2017a WRK4). The N-terminus has the median length of 389 amino acids and partially resembles the SEFIR family ([78, 79]) of TIR clan (HHblits hit probability of 90%). The TIR domain was reported in NLRs from plants, bacteria and Chytridiomycota [80, 20, 5]. The NLR proteins in this cluster are often associated with C-terminal repeats of Ankyrin, HEAT and WD40 types. In addition to the NACHT-based architectures present in the cluster, the web-based profile HMM search revealed several additional homologs in NLRs with the NB-ARC TPR domains. Interestingly, homologous domains are also present as separate proteins in Mucormycota *Rhizophagus irregularis*, a species related to *Mortierella*, and in association with NACHT WD40 and NACHT HEAT in *Mycoavidus cysteinexigens*, in accordance with the possibility of horizontal gene transfer [81]. A good quality structural model supports homology to TIR and HET domains (Fig. S2c).

## Amyloid-like motifs in short N-termini

The largest deficiency in the annotation coverage concerns short N-terminal sequences (length below 150 amino acids). Only a few percent of them (242 out of 3441) was annotated as so called prion-forming domains (PFD) [32, 14], consisting of the known fungal amyloid signaling sequences (FASS). However, many other fungal NLRs include short N-termini with a stretch of amino acids similar to already known signaling amyloids, which are characterized by a hydrophobic pattern typical to beta-sheets and frequent asparagines, glutamines and glycines. Moreover, the repertoire of already described FASS, merely three, is significantly smaller in comparison to bacterial amyloid signaling sequences (BASS) of which 10 families were recently identified in association to NLR proteins and the Bell domain [20]. Thus, in search for potential additional fungal amyloid signaling motifs we decided to thoroughly scan the short N-termini of NLRs with a probabilistic grammatical model inferred from ten families of bacterial amyloid signaling motifs (BASS) and shown to be sensitive to fungal amyloid signaling motifs [43].

In total, N-termini of 3441 sequences from 54 clusters with average sequence length up to around 150 amino acids were analyzed. Very high scoring fragments (see Methods) were found in 18 clusters with 1456 sequences. This included all 8 clusters (including 592 sequences) with at least one PFD-like annotation. The N-terminal sequences were made non-redundant at the identity level of 90% and submitted to motif extraction with MEME [82]. A total of 51 motifs were found for the E-value threshold of 1, for which profile HMMs were built in a two-stage procedure as in [20] (see Methods). Sequences of the 16 motifs were consistent with the grammatical model of BASS. These candidate amyloid-like motifs are presented in Fig. 3a. Eventually, their HMM profiles were used to scan all NLR N-termini at least 10 amino-acids long, comprising also sequences not included in the 127 clusters with 20 or more members (Tab. 3).

Not surprisingly, some of the 16 motifs clearly corresponded to the three FASS families: HRAM (NLR13, found in 131 sequences), PP (NLR07, 296),  $\sigma$  (NLR28, 71). The overall recall of 498 hits is twice higher in comparison to the combined Pfam-based approaches (242). Several hits of another two motifs, NLR12 and NLR40, overlapped with the NLR13 (HRAM) matches (Fig. 3b). Moreover, the HMM scan with a generalized HRAM profile based on HRAM dataset from [33] recognized 1/3 of NLR12 and 2/3 of NLR40 motifs at E-value of 0.01, thus indicating that these two

classes are related to HRAM. Indeed, the NLR12 motif (Fig. 3a) is apparently similar to HRAM3 [33]. In addition, the G-hydrophobic-Q-hydrophobic-G pattern of NLR39 motif resembles NLR07 (PP). Five other motifs (NLR17, NLR19, NLR20, NLR32 and NLR34, in 138 sequences altogether) were difficult to assign to the known families. The final and the largest subgroup (689 sequences) consisted of five motifs (NLR05, NLR08, NLR22, NLR29 and NLR44) with hits substantially overlapping NLR22 hits. This large group is specific to basidiomycetes except of a dozen of NLR22 hits overlapping ascomycotal NLR28 ( $\sigma$ ) (Fig. 3b). While most motifs are distributed in larger taxonomic branches, two motifs are more restricted: NLR17 is specific to *Amanita muscaria* (strain Koide) and NLR19 to genus *Tuber*. A combined NLR19 + NLR34 configuration was found in five highly homologous sequences from *Tuber melanosporum* (Fig. 3b).

Significant numbers of similar motifs in C-termini (100aa) of genomically neighboring (20kbp) proteins were found only for motifs representing the three FASS families (NLR07 in 37 sequences, NLR12 in 4, NLR13 in 16, and NLR28 in 42) and for NLR32 (11 sequences). This suggests that NLR32 defines a new family of amyloid signaling motifs. (For further computational and experimental verification, see below.)

Notably, the amyloid-like motifs differ in their position in NLR N-termini. While motifs from the NLR05/NLR22 group are usually situated in the very terminus, most HRAMs (NLR12/13/40) and PPs (NLR07/39) are located at positions 5–9. Moreover, NLR32 and  $\sigma$  motifs (NLR28) are shifted further C-terminally with relative majority at positions 20–49 and 50–99, respectively (Fig. 3c). In addition, a couple of dozens of amyloid-like motifs of various families (including 17 NLR05 and 7 NLR07) were found located centrally or C-terminally in longer N-termini. Some of them form combined architectures with annotated domains, most notably with NLR\_PRDR (NLR05 in 10 sequences from *A.bisporus*) and MLKL-likes (5 BaMLKL + NLR05 in *Laccaria bicolor*, 4 HeLo-like + NLR28 and 1 HeLo-like + NLR07 in various Ascomycota).

Regardless of the evidence of amyloid signaling, all 16 motifs are likely to assume the beta-arch fold typical to known FASS and BASS as from 45 to 95% motif instances pass the fold prediction threshold of ArchCandy (0.56). The only exceptions are two shortest motifs NLR05 (28%) and NLR44 (none), probably because they comprise only parts of the actual amyloid-like motif (Fig. 3a). For three motif profiles matches were found in the PDB database [83], namely pdb:3erb [84] for NLR05/NLR22, and pdb:4q2w:A [85] for NLR34. The both solved proteins are hydrolases and the matched fragments possess a beta-arch-like fold.

## **A reverse approach: amyloid-like motifs in C-termini of effector proteins**

In order to complement the search for amyloid signaling motifs in NLRs and verify discovery of the fourth NLR-related FASS family, we adapted the approach recently used for identification of 10 families of BASS in NLR-related proteins in bacteria [20].

We iteratively searched for remote homologs (using HMMER [86]) of effector domains related to NLR proteins, starting with 19 Pfam profiles of N-terminal domains of NLRs reported in [5] (see Methods). We found almost 140,000 sequences with the single domain architecture and C-terminus between 10 to 150 amino acids in length. For each domain family separately, the C-termini were made non-redundant at identity of 70% (using CD-HIT [87, 88]) and searched for short (10–30 amino acids) motifs using MEME [82]. This resulted in around 800 motifs, for which profile HMMs were built, as in the previous section (see Methods). In addition, we extracted all fungal NACHT and NB-ARC proteins (according to Pfam) with N-termini between 10 and 150

amino acids. The N-termini were then searched for MEME motifs, for which, eventually, 49 profile HMMs were generated. NLR N-termini and effector C-termini were scanned with the combined set of effector- and NLR-side motifs, and the hits in both groups of proteins (at E-value of  $1e-2$ ) were matched based on genomic proximity (up to 20kbp) of genes encoding the proteins. Eventually, for 19 motifs, at least 3 non-redundant pairs of motif instances were found.

These motifs were clustered on the basis of their co-occurrence in pairs of sequences, with three clusters corresponding to the already known families PP,  $\sigma$ , and HRAM (Fig. S3). Two additional motifs with few pairs apparently resemble HRAM2 and HRAM4 [33], respectively, and one another motif resembles the C-terminal part of the  $\sigma$  family motifs (Fig. S3). The fourth largest family of motifs exhibits a distinctive conserved pattern FxGxGxQxxGxGxF, which clearly corresponds to the NLR32 motif in Fig. 3. Since in both searches the motif was found associated uniquely with the PNP\_UDP domain, we term it PUASM, or the Pnp\_Udp-associated Amyloid Signaling Motif. The NLRs with the PUASM motif proteins are annotated either as NACHT or NACHT WD40. All matched instances of the PUASM motif come from various *Pezizomycotina* species. Finally, we found one more distinct motif related to PNP\_UDP, however only present in four pairs (PF01048\_040 in Fig. S3).

Overall, the ASM differ in type of associated effector domain, either pore-forming (HeLo and HeLo-like for HRAM/NLR13), enzymatic (PNP\_UDP for HRAM/NLR12, NLR32 and PF01048\_040), or both (PP/NLR07 and  $\sigma$ /NLR28). Interestingly, while the NLR13 motif was typically found as a double in C-termini of HeLo and HeLo-like domains, for the second HRAM-like, NLR12, only single instances were found in C-termini of PNP\_UDP\_1 effector proteins. This suggests a different mode of operation despite their similar sequence profiles.

To check the possibility that proteins cooperating through the amyloid signaling are encoded by sequentially distant genes, we analyzed co-occurrence of particular amyloid-like motifs in N-termini of NLRs and C-termini of previously described effector domains [5] in entire genomes. Non-singular C-terminal hits and genomic co-occurrences were found only for FASS and NLR32 (Tab. 3).

## Amyloid-like motifs in Basidiomycota

With the NLR-related amyloid signaling previously described in multicellular bacteria and Ascomycota, apparent is the lack of evidence of this mechanism in Basidiomycota. On the other hand, we found numerous homologs of the pore-forming HeLo and HeLo-like domains in Basidiomycotal NLRs. Thus, we used them for searching the entire Basidiomycota genomes for homologs separate from NLR domains (web-based profile HMM search with standard parameters). We identified hundreds of such putative singular pore-forming domains, which — because of their potential to cause the cell death — can be expected to be under control of other proteins. As in Ascomycota such control is exerted by NLRs through the amyloid signaling motifs, we scanned the identified MLKL\_NTD homologs against ASM motif profiles and grammars. However, motifs resembling ASMs were identified only in a few out of 500 sequences and no associated motifs were found in the neighboring NLRs. Yet in two cases the motif pairs occurred when entire genomes were considered (Fig. 4a). In *Moniliophthora roreri* (strain MCA 2997) there is a 18 amino-acid long motif apparently shared between two MLKL\_NTD-like C-termini and 26 short NACHT N-termini (Fig. S4). In addition in *Fibularhizoctonia* sp. CBS 109695 there is a conserved pattern shared between two MLKL\_NTD-like C-termini, eight short NLR N-termini (including KZP25847 with NLR20 in-

stance) and additional five NLR proteins where the pattern is situated between MLKL\_NTD-like and NACHT domains (including KZP30127 and KZP3012 with NLR22 instances), see alignment in Fig. S5. It would suggest a possibility that in *Fibularhizoctonia* proteins with the N-terminal and C-terminal amyloid-like motifs are pseudogenes, especially that three NLRs in this group are untypically short (less than 200 amino acids). However, NLRs with N-terminal and mid-sequence ASMs differ in domain configuration with the former belonging to NACHT, NACHT ANK and NACHT VHS architectures, while the latter are all of the NACHT TPR type (Fig. 4a). In turn, for *M. rorei*, we found only one protein with the MLKL\_NTD + AAA\_16 architecture (ESK90106.1) and the linker sequence between the domains does not resemble an amyloid-like motif.

In addition, we investigated two Agaricomycetes species with proteins comprising of a singular HeLo domain and a C-terminal double HET-s motif. In the genome of *Sphaerobolus stellatus* (strain SS14), which includes four such C-termini, we found at least eight NACHT NLRs with N-termini comprising of single HRAM-like motifs (Supp. Fig. S6). Two instances (in KIJ28522 and KIJ30800) are recognized as the NLR13 HRAM motif at E-value around 0.01. This strain is the only case where an NLR and three HeLo proteins are situated on a single contig in genome assembly (Fig. 4bc). Interestingly the shortest distance between genes encoding NLR and HeLo is 95 kbp. The second species, *Gymnopus luxurians* (strain FD-317 M1), includes one protein with HeLo + double HET-s motif architecture. While we did not find any typical HRAMs in N-termini of 200 NLRs, several dozens included an instance of the NLR05/08/22/44 motif meta-family. Eventually we extracted N-terminal motif with highest scores according to the PCFG model and aligned them with Mafft, which revealed a 25-residue long core pattern. Interestingly, the alignment exhibited features characteristic to HRAMs: the N-terminal pattern of three hydrophobic residues and the C-terminal G[DN] bigram (Supp. Fig. S7). The 32 motifs are associated with NB-ARC, NACHT, NACHT WD and NACHT TPR domain architectures. Taken together these analyses strongly suggest that the NLR-associated amyloid signaling process also occurs in Basidiomycota.

## Experimental validation of a novel amyloid signaling motif

The alignment of the PUASM motif pairs (Fig. 5a) reveals high similarity of PNP\_UDP- and NLR-side sequences in the core region covered with the NANBNtm\_035 pattern. Some divergence can be seen C-terminally, with GND pattern prevailing in PNP\_UDP-side motifs, while ARD pattern — in NLR-side motifs. Interestingly, these 3-mers can be found in C-termini of already known amyloid signaling motifs HRAM1 [33] and BASS2 [20], respectively. Further four residues of the C-terminal extension of the motif exhibit a hydrophobic pattern well conserved in pairwise alignments (Fig. 5a). On the other side, N-terminal extensions of the PUASM profile matches often include histidine in the PNP\_UDP side and glutamic acid in the NLR-side. This, together with the overall composition of the N-terminal extensions, suggests some role of the charge complementarity.

To check if biochemical properties of the PUASM motif are consistent with its presumed role as the amyloid signaling motif, we experimentally analyzed a representative pair of motifs of this family, namely, PNP\_UDP-side C-terminal EQB50682.1\_332\_355 (VFHGKGIQHTGSGNFSVGNLDSIS) and NLR-side N-terminal EQB50683.1\_9\_31 (FHGHGIALSGAGNITVGGDFIIG) from a plant pathogenic fungus *Colletotrichum gloeosporioides* Cg-14 [89] (Tab. S1 and Fig. S8). The selected fragments entirely cover the matches of PUASM profiles and the pairwise conserved C-terminal extensions.

***In vitro* study of amyloidogenic properties.** The aggregation propensities of the PUASM peptides were determined experimentally using the Attenuated Total Reflectance – Fourier Transform InfraRed spectroscopy (ATR-FTIR), Atomic Force Microscopy (AFM), and the Thioflavin T fluorescence assay (ThT), in accordance to the MIRRAGGE standard [90]. The ATR-FTIR spectroscopy allows determination of secondary structure and monitoring structural changes of peptides upon aggregation processes [91, 92, 93], while AFM is useful for detection and visualization of aggregates [94]. In turn, ThT is considered to be a “gold standard” for identifying amyloid fibrils [95, 96]. It is widely accepted that a combination of these techniques is necessary to ascertain if a particular peptide or protein is able to form the amyloid assemblies [97, 98, 90].

Analysis of the ATR-FTIR spectra in the range of 1750–1500  $\text{cm}^{-1}$  (Fig. 5bcde, S9 and Tab. S2, S3) confirmed aggregation properties of studied peptides. The maximum of Amide I' band was observed at 1625  $\text{cm}^{-1}$  and 1630  $\text{cm}^{-1}$  for EQB50682.1\_332\_355 and EQB50683.1\_9\_31, respectively. This signature is considered to be a spectroscopic marker of the formation of intermolecular aggregates [99] and corresponds to the cross- $\beta$  amyloid architecture [92]. High absorbances in region of 1670–1660  $\text{cm}^{-1}$  are characteristic of parallel  $\beta$ -helix structure and commonly observed in infrared spectra of peptides and proteins with beta solenoid conformations, e.g. HET-s [100] or PrPSc [101]. Further analysis of the deconvoluted and derivative spectra revealed more detailed information about the structure of aggregates. While for both studied peptides the aggregation process was observed immediately after dissolving, N-terminal EQB50683.1\_9\_31 aggregated faster and formed more rigid assemblies (Fig. 5bd). The percentage area of the subband corresponding to intramolecular  $\beta$ -structure at about 1630  $\text{cm}^{-1}$  was 31% and 14%, for peptide EQB50683.1\_9\_31 and EQB50682.1\_332\_355 respectively. Moreover, the low-frequency component of Amide I' band, which is generally assigned to highly ordered intermolecular structures, in the spectra of peptide EQB50683.1\_9\_31 is observed at 1616  $\text{cm}^{-1}$ . In turns EQB50682.1\_332\_35 exhibits this signature at higher wavenumbers (1616  $\text{cm}^{-1}$ ), indicating a looser fibrillar structure [102].

Atomic Force Microscopy (AFM) images of both PUASM peptides were acquired for two conditions related to the spectroscopy studies: after dissolving, and after 40 days of incubation in room temperature. The aggregation process of the peptides was present already in the sample after dissolving as the fibers with height of  $3.44 \pm 0.3$  nm and  $3.33 \pm 0.3$  nm, respectively for EQB50682.1\_332\_355 and EQB50683.1\_9\_31, were observed (Fig. 5fg). The height of the object observable in AFM is comparable with the size of the HET-s peptides obtained by the solid-state NMR technique (pdb:2kj3) [30]. Peptide aggregation was further enhanced in the samples imaged after 40 days (Fig. 5hi) when the height of the aggregates reached  $10.08 \pm 0.9$  nm and  $14.28 \pm 1.3$  nm, respectively for EQB50682.1\_332\_355 and EQB50683.1\_9\_31. Along with the increasing aggregation process, the fibers underwent morphological changes. In the case of peptide EQB50682.1\_332\_355, greater flexibility of the fibers was observed, which may be related to the looser packing of the fiber structure. On the other hand, in the case of peptide EQB50683.1\_9\_31, a greater ability to form aggregates and greater stiffness of the fibers were observed, which may be related to tighter packing of the fiber structure. These observations are in line with the ATR-FTIR measurements (Fig. 5ce).

Thioflavin T (ThT) fluorescence assay was used to assess the kinetics of aggregation process. We observed a typical sigmoidal nucleation–polymerization curve for peptide C-terminal EQB50682.1\_332\_355, starting with a lag phase of 2 hours (Fig. S10), followed by a rapid growth phase from 2–2.20 h, and ending at a stable plateau with the maximum ThT intensity. A signifi-

cant increase in the fluorescence emission was observed for peptide N-terminal EQB50683.1\_9\_31 (about 5 times higher than for peptide EQB50682.1\_332\_355). What is more, the lag phase was not observed, the curve starts from the elongation phase, which corresponds to the oligomers and protofibrils states. The steep ThT curve with quicker attainment of plateau indicate faster aggregation process of peptide EQB50683.1\_9\_31 in comparison to peptide EQB50682.1\_332\_355.

***In vivo* study of signaling capability.** It was previously reported that fungal, bacterial and mammalian amyloid motifs could form prions *in vivo* in the *Podospora anserina* model [27, 20, 103, 36]. To determine if PUASMs could also form prions *in vivo*, we expressed the PNP\_UDP-side C-terminal EQB50682.1\_332\_355 (VFHGKGIQHTGSGNFSVGNLDSIS) from a plant pathogenic fungus *Colletotrichum gloeosporioides* Cg-14 [89] in *P. anserina*. Prion propagation and transmission was monitored using fluorescence microscopy as the formation of cytoplasmic fluorescent foci. The GFP-PUASM fusion lead to an initial diffused fluorescence signal (Fig. 6). Upon subculture, the number of transformants showing cytoplasmic foci gradually increased over time as typically observed for other prion amyloid motifs [103]. The rate of this spontaneous transition from the diffuse state to the aggregated state was monitored over 75 days (Tab. S4). In contrast to the GFP-PUASM construct, fusion constructs displaying the motif N-terminally (PUASM-GFP and PUASM-RFP) remained diffused and did not form foci. A similar situation was observed previously for the HELLF and RHIM motifs for which N-terminal position of the GFP/RFP inhibited foci formation [103].

For all the amyloid signaling motifs tested so far in *P. anserina*, the aggregated state could propagate as a prion by cytoplasmic contact with a strain already expressing the fusion protein in the aggregated state. This assay, referred to as induced prion formation, was also performed for the PUASM fusion proteins. The GFP-PUASM fusion was converted to the foci state with high efficiency after cytoplasmic contact with a donor strain containing foci (Tab. S4, Fig. 6). Again, for the PUASM-GFP and PUASM-RFP proteins prion conversion was not observed. We conclude from these experiments that the GFP-PUASM fusion protein behaves as a prion *in vivo* in the *Podospora* model. The spontaneous and induced prion conversion was however less efficient than for other amyloid motifs [20, 103, 36].

## Discussion

**N-terminal annotations of fungal NLRs are not evenly distributed.** In previous studies we computationally screened N-terminal domains of fungal NLRs using profile Hidden Markov Models (pHMM) from the Pfam database directly and complemented the search with several Pfam-like inhouse models [14, 5]. Here we expanded the most recent analysis with a more sensitive search using the state-of-the-art clustering offered by MMseqs2 and pHMM-pHMM searches with HHblits. The study increased the overall Pfam annotation coverage of N-terminal domains by about 16% (or 19% when MLKL\_NTD from CDD is counted), but also highlighted remarkable deficiencies in availability of annotations.

While the vast majority of longer domains is at least partially annotated, this is true only for a definite minority of shorter domains (Fig. 1c). As about 3 400 short domains were assigned to clusters made with at least 20 sequences, the shortage of annotations cannot be easily explained by the lack of conserved sequential features. Instead, one of the reasons is the pHMM model itself, which

by assessing each alignment position independently (except for indels) is not statistically powerful enough when dealing with short sequences. In other words, pHMM models of more diverse families of short sequence fragments (e.g. 20–40 amino-acids long) cannot be sensitive and specific at the same time [43]. Currently, the problem can be at least partially addressed by using more complex and computationally demanding protein sequence models, such as probabilistic context-free grammars (PCFG) [104, 20, 43] and co-evolutionary Potts models [105, 106, 107]. Another viable option are the recurrent and attention-based neural networks, which have enough computational power to describe relevant dependencies in protein sequences [108, 109, 110]. However, while modern neural networks have been successfully applied to annotation of protein families [111, 112], their performance in modeling short protein sequence fragments is yet to be evaluated.

**Differences between the Asco- and Basidiomycota effectors.** The Pfam annotation coverage strongly depends on taxonomic scope. Specifically, the Basidiomycota are covered three times less than Ascomycota (20–25% vs. 70%, Fig. 1d), with other fungi in between (about 40%). Moreover, the disproportion in a share of annotated NLR effectors in asco- and basidiomycetes holds also when comparing N-terminal extension in the same length ranges. In addition, domains with homologs in other Eukaryota were significantly better annotated than domains specific to fungi (80% vs. 20%, Fig. 1e). While better coverage of more universally spread domains is not surprising, taken together, our results highlight the fact that the NLRs of fungi, and especially Basidiomycota, are still not sufficiently described. Certainly, this may have implications for understanding the NLR system in general, e.g. regarding its evolutionary history [15, 26].

The previous and present annotation reveals some resemblances between the effector domains found in ascomycetes and basidiomycetes, the most notable being the high proportion of domains in the Goodbye/HeLo/MLKL group (Fig. S1). Yet, some marked differences also come to attention, like the lack of occurrence of HET or PUP domain basidiomycetes (although these domains occur in other (non-NLR) domain architectures). Another difference is the length distribution of the effector domains that globally appear much shorted in Basidiomycota. What are the underlying causes of these marked differences in the evolutionary history of the NLR-family in these two major fungal groups remains at present elusive.

**MLKL-related HeLo- and Goodbye-likes form the most abundant superfamily of effectors.**

Analyzing the annotation results, we found clusters with overlapping HeLo/HeLo-like and HeLo-like/Goodbye-like hits assigned through the HHblits search. The HeLo domain is a fungal homolog of human MLKL, plant RPW8 and bacterial Bell domains, while HeLo-like is closely related [32, 40, 20]. It is understood that upon oligomerization, these domains, whose central part is a four-helix bundle, expel a N-terminal alpha-helix to form a pore targeting the membrane and thus induce the cell death. The structure and function of Goodbye-like is yet to be established experimentally. The HeLo-like/Goodbye-like cases, typical to Basidiomycota, were usually below the default threshold for direct Pfam annotation but were often assigned the MLKL\_NTD annotation of the Conserved Domain Database. The Goodbye/HeLo/MLKL-like superfamily comprises of approximately 30% of all N-termini of fungal NLRs (Tab. 1) and is the most numerous both in Asco- and Basidiomycota.

Both the distribution of associated nucleotide-binding domain and C-terminal domains, and the paralog-to-ortholog ratio for Goodbye-like and HeLo-like domains are similar [14], which may

suggest similarities in their mode of operation. However, Goodbye-likes in NLR N-termini are often associated with another annotated effector domains, which is untypical for HeLo-likes (Tab. 1). Moreover, the opposite is true for association with the amyloid signaling motifs, which is common to HeLos, HeLo-likes and basidiomycotal MLKL\_NTDs but not to Goodbye-likes. The multiple sequence alignment and predicted structures of Goodbye/HeLo/MLKL-like domains reveal the matching region covers the four-helix bundle (Fig. 2). In a plant homolog of HeLo, the N-terminal helix of the bundle (and entire protein) is known to play a significant role in triggering the cell death process [72]. However, in BaMLKL and Goodbye-likes, the bundle is extended N-terminally by one or more helices. Thus, while the common evolutionary ancestry of HeLo-like and Goodbye-like seems rather evident, the question of their functional similarity remains open.

**Effector families distribution: the widespread and the confined.** In addition to the Goodbye/HeLo/MLKL-likes, the other two superfamilies highly abundant in Ascomycota (a dozen of percent each, Fig. 1e and Fig. S1) are SesB-like AB hydrolases and the purine and uridine phosphorylases, which both provides enzymatic functions. Apart from them, the only other more frequent longer domains are Patatin (2-3% of Asco- and Basidiomycota) and HET (ca. 3% of Ascomycota). Interestingly, a couple of dozens of domains from the TIR (Toll/Interleukin-1 receptor-like) clan, remotely homologous to HET, were found in Pezizomycotina and in Chytridiomycota. All other longer domains — whether determined by the Pfam annotation or the MMseqs2 clustering — were represented by less than 100 sequences. This indicates that the current Pfam annotations (plus MLKL\_NTD and a few inhouse profiles) cover all widely spread abundant domains. At the same time, there seems to exist a substantially large corpus of thousands of specialized N-termini, sometimes confined to narrow taxonomic branches. While some of them may be formed with a tuple of known domains, other could represent novel families (likely being difficult targets for structure prediction due to small alignments). Naturally, we cannot rule out that in some cases larger families were superficially partitioned into small clusters (with less than 20 members). Nevertheless, with regard to our previous analyses [14, 5], this study suggests less diversity in major effector classes (5-7 rather than 12-13), but highlights a likely abundance of specialized domains.

**A large fraction of the effector domain are tentatively involved in regulated cell death.** With the limitation that the evolutionary relation of Goodbye-like domains does not necessarily imply functional similarity, it appears that a substantial fraction of the effector domains in both ascomycetes and basidiomycetes is predicted to control regulated cell death. Involvement in regulated cell death has been reported not only for the HeLo/Goodbye/MLKL group but also for the HET domain [113], the Patatin [18] domain and more indirectly for the SesB-like domain [40]. One needs to add to this list the amyloid signaling motifs that control separate downstream cell-death effector domains. Globally, it would appear of at least one third to one half of the fungal NLRs could be involved in some kind of regulated cell death process. This high proportion raises the question of whether some of the other domains (whether annotated or not) could also play a role in regulated cell death.

**NLR-associated amyloid-like motifs in Ascomycota are more diverse than previously reported.** We significantly improved the annotation coverage of NLR amyloid signaling domains. By pipelining clustering with MMseqs2, filtering sequences with our recent PCFG-based generalized model

of beta-arch amyloid motifs, extracting specialized motifs with MEME, and generating their profiles with HMMER, we identified amyloid-like motifs in roughly one third of the 3 400 short N-terminal domains (Fig. 1 and Tab. 3).

In Ascomycota, within several hundreds of extracted motif instances, we discovered a new motif family, NLR32, which is uniquely associated with the PNP\_UDP domain (thus termed PUASM, Fig. 3). Similarly to other ascomycotal amyloid signaling motifs, HRAM, PP and  $\sigma$ , effector proteins with C-terminal PUASM are often coded by direct genomic neighbors of the PUASM–NLR genes. Such genomic co-localization facilitates co-inheritance of the two genes of the functional unit in the event of a recombination process. This may be of special importance for the NLR signaling pathway, which is polymorphic in population given the death-and-birth evolution. The amyloid properties of the NLR32 motif were confirmed experimentally using a representative pair of N- and C-terminal sequences. While both of them generated amyloid-like fibers, the N-terminal motif aggregated quicker and formed more rigid structures, which is consistent with its expected role as the template for aggregation of the C-terminal motif. (In depth study of the co-aggregation process is left for a separate study.) The effector-side PUASM sequence was shown to be capable of forming prions *in vivo* in the *Podospora anserina* model.

Interestingly, we found the PNP\_UDP domain associated also with a HRAM motif variant (NLR12) and with one another amyloid-like motif discovered in this study (PF01048\_40, Fig. S3).

For NLR12, only in a few cases the effector–motif and motif–NLR pairs present in the genome were co-localized (that is encoded by adjacent genes). In some genomes, the NLR-side motifs were relatively more frequent (mean ratio 5.7:1, Tab. 3). We speculate that the presence of many NLRs controlling the same effector could potentially relieve the need for genomic co-localization.

**Less diversity of ASM in fungi as compared to bacteria even after extensive search.** Although, we have successfully identified new motifs and validated the most abundant of them, the expand diversity of ASM remains higher in bacteria than in fungi using similar identification procedures, which is not inconsistent of the larger phylogenetic breath of the scanned bacterial genome as compared to the fungal ensemble.

**NLR-related amyloid signaling is present in Agaricomycetes.** In Basidiomycota, virtually none of the hundreds of NLR-side amyloid-like motifs found in our survey was genomically co-localized with C-terminal side motifs. Moreover, motifs closely resembling NLR-side amyloid-like motifs were absent in C-termini of previously identified effector proteins whatever their location in genome. Eventually, by analyzing C-termini of single-domain MLKL-like proteins we identified likely associations in *Moniliophthora roreri* and *Fibularhizoctonia* sp. CBS 109695. Similarly by analyzing two genomes with HeLo-HRAM-HRAM architectures, we found HRAM-NLR proteins in *Sphaerobolus stellatus* and NLR05/NLR22-NLR proteins in *Gymnopus luxurians*. As NLR05/NLR22 meta-family of amyloid-like motifs shares with HRAM the hydrophobic pattern and the largely conserved C-terminal G, the study suggests that the amyloid signaling could be preserved despite substantial sequence variability. Eventually, limits of this preservation requires experimental verification [36, 114].

Taken together, presented results support the presence of amyloid signaling in Basidiomycota, or more specifically in Agaricomycetes, in the context of NLR-based regulation of HeLo-/MLKL\_NTD-likes. Moreover, they suggest that NLR05/08/22/29/44 meta-family of motifs is a

basidiomycotal variety of the HRAM motif or its homolog. However, there are significant differences with regard to Ascomycota. First, while NLR-side amyloid signaling motifs are present in roughly half of Ascomycota strains, they were only found in 1/4 (30%) of Basidiomycota (Agaricomycetes) strains. Second, while there are typically only few such motifs per ascomycotal strain, there are usually dozens per basidiomycotal strain. At the same time, basidiomycotal effector-side C-terminal motifs are seemingly less frequent than NLR-side N-terminal motifs (Fig. S4-S7). Indeed, the high number of NLR-side motifs corresponds to enrichment of basidiomycotal sequences among shorter N-terminal domains (Fig. 1b). This could at least partially explain the lack of genomic co-localization of NLRs and their effectors linked by amyloid signaling motifs.

One interesting finding is presence NLRs with intra-proteins amyloid-like motifs in *Fibularhi-zoetonia* sp. CBS 109695. Different central and C-terminal domain association in comparison to NLRs with N-terminal ASM-likes suggest also different functions of the motifs in both cases. Therefore, we hypothesize that these internal amyloid motifs may serve as scaffolds to stabilize the NLR oligomers, similar to cRHIM motifs in the RIP1K/RIP3K complex [45]. In these lines, it is possible that also some other amyloid-like motifs identified in the current study but with no matching effector-side counterparts participate in the assembly of the NLR signalosome or are involved in interactions with motifs located outside the C-terminus of the associated protein.

## Materials and Methods

### Computational methods

**Annotation of NLR N-termini.** A set of 36,141 NLR proteins from 487 fungal strains was identified in a previous study through the PSI-BLAST [115] search among completely sequenced fungal genomes in the NCBI nr database [14, 20]. 32 962 N-termini at least 20 amino-acids long (91%), delimited according to the NACHT or NB-ARC query matches, were further considered, of which 18,674 (57%) were annotated using direct matches to Pfam [25] or inhouse HMM profiles (Supplementary File 2) [14, 20]. The set of N-termini at least 20 amino-acid long was clustered with MM-seqs2 [46] in mode 1 (21 758 N-termini in 127 clusters, 15 105 already annotated). Then, sequences in each cluster with at least 20 members were aligned using Clustal-Omega [116] (Supplementary File 3) and searched for homologs in UniRef30 [47, 48] using HHblits [49] (parameters: `-e 0.001 -n 2 -E 0.01 -Z 1000000 -M 50`). Subsequently, the resulting alignment was used to search Pfam (HHblits parameters: `-e 0.001 -n 1 -E 1 -Z 1000000`). The clustering required mutual coverage of at least 80% of sequence length, and the annotations were only assigned to sequences which covered at least 50% of the match to the Pfam profile. The resulting cluster-level annotations were retained only if the alignment match to the Pfam profile covered at least 50% of the profile length, and assigned only to individual sequences which covered at least 50% of the match. After completing the main processing, the set of N-termini was re-scanned for the Crinkler domain (PF20147) added recently to the Pfam database.

The tabularized results of the annotation are provided in Supplementary File 4. The overlapping Pfam annotations were resolved as in [14, 20]. The double HeLo/HeLo-like annotations were kept in the Supplementary File 4 and in Tab. 1 but were represented as HeLo in Fig. 1f and Fig. S1. In addition, basidiomycotal sequences from clusters doubly annotated as Goodbye-like/Helo-like, as well as from clusters with CDD [65] MLKL\_NTD annotations, were denoted as BaMLKL (see

Results).

**Comparative analysis of Goodbye-, HeLo- and MLKL-likes.** The largest clusters annotated as HeLo, HeLo-like, Goodbye-like and BaMLKL were submitted to AlphaFold2 structure prediction [69] through the ColabFold advanced notebook [70]. Standard parameters of the notebook were applied except of (1) using the cluster alignments instead of searching genetic databases, (2) trimming off fragments just upstream the NACHT domain were applicable. Successful models – with the mean predicted IDDT score [69] above 70 overall, and around 80 or more for the core helix bundle – and respective ColabFold outputs are provided in Supplementary File 5. For each cluster, the highest rank model was selected and structurally aligned to the experimentally solved MLKL domain (pdb:6zvo) using TM-align [71].

**Characterization of unannotated longer N-termini.** For five unannotated clusters with at least ten members at the identity threshold of 70% and the median length above 100 amino acids homologs were searched in UniProt [117] through the web-based hmmsearch with standard parameters [118], and prediction of the three dimensional structure was attempted using AlphaFold2 [69] through the ColabFold advanced notebook [70]. Standard parameters were used except of adding the MMseqs2 alignments to input (sequences just upstream the NACHT domain was trimmed off). Good quality structures (the predicted IDDT score above 70) were obtained for three clusters, KEY84097, KFH66451 and PQE30996 (Supplementary File 6). The proposed annotations for the five clusters (Tab. 2) are assigned to member sequences in the table in Supplementary File 4 and included in the TIR-like and “other” groups in Fig. 1f and Fig. S1.

**Extraction of amyloid-like motifs in short N-termini.** A subset of 54 NLR N-termini clusters with mean/medium sequence length at most 160/161 amino acids was scanned using the PCFG-CM software [119, 43] probabilistic grammatical model inferred from ten BASS families [43] (Supplementary File 7) with scanning window of 20 to 40 amino acids and the smoothing factor of 10 PAM [43]. Very high scoring fragments (maximum log<sub>10</sub> score at least 3.5, mean log<sub>10</sub> score above 1.67) were found in 18 clusters with 1456 sequences. The N-terminal sequences were made non-redundant at the identity level of 90% using CD-HIT 4.7 [87, 88] and submitted to motif extraction with MEME 5.0.5 [120, 82]. For each of 51 motifs found at the E-value threshold of 1, HMM profiles were built with HHMER 3.2.1 [86] and used for searching against the full set of grammar-fitting N-terminals (at E-value of  $1e - 2$ ). Then, obtained hits were extended by 5 amino acids in each direction and realigned using Clustal-Omega with the auto parameter. For each motif, the extended sequences were re-examined for consistency with the grammatical model (maximum log<sub>10</sub> score at least 3, mean log<sub>10</sub> score above 1). For 16 motifs which passed the grammatical filter, the alignments were used to build final HMM profiles (Supplementary File 8).

**Analysis of N-terminal amyloid-like motifs.** The HMM profiles of the 16 motifs were used for scanning all N-termini longer than 10 amino acids (the independent domain E-value threshold of  $1e - 2$ ). Hits are included in the Supplementary Tab. 4 with coordinates (outermost in rare cases of double ASM hits). For further analysis only hits in N-termini shorter than 200 amino acids not located beyond position 150 were considered. Motif sequences in envelopes of 5 amino acids were

tested for the beta-arch structure with ArchCandy 2.0 [42] using the recommended threshold of 0.56.

For each motif-containing NLR sequence, proteins coded by genes within the  $\pm 20$  kbp neighborhood of the genes encoding these NLRs were fetched from NCBI GenBank [121] or EMBL ENA [122] using an in-house Python (version 3.7.3) script (aided by requests [123] and xmldict [124] packages). The set was confined to proteins in the length range of 200-400 amino acids, which is typical for proteins with single domain architectures known to be associated to NLRs via amyloid signaling [33, 20]. Next, C-termini (100 amino acids) of the found neighboring proteins were scanned for the presence of the motifs using HMMER (the independent domain E-value threshold of  $1e - 2$ ). Pairwise hits of the same motifs in N-termini of NLRs and C-termini of genomically neighboring proteins are collected in Supplementary File 9.

**Homology search of effector domains.** Homologs of NLR effector domains, as listed in [5], were iteratively searched using Pfam profiles: Pkinase (PF00069), Peptidase\_S8 (PF00082), C2 (PF00168), PNP\_UDP\_1 (PF01048), TIR (PF01582), Patatin (PF01734), RelA\_SpoT (PF04607), DUF676 (PF05057), HET (PF06985), PK\_Tyr\_Ser-Thr (PF07714), PGAP1 (PF07819), Abhydrolase\_6 (PF12697), CHAT (PF12770), TIR\_2 (PF13676), HeLo (PF14479), NACHT\_N (PF17100), SesA (PF17107), Goodbye (PF17109) and Helo\_like\_N (PF17111). First, HMM profiles for each of the domains were used for searching against a local copy of the non-redundant protein sequences database (NCBI's "nr", downloaded in November 2019) [121] using HMMER 3.2.1 with E-value of  $1e - 2$ . Found proteins were then used to build the new HMM profiles and the search was repeated until the number of hits did not change by more than 7%. In addition, all fungal NACHT (PF05729) and NB-ARC (PF00931) proteins (as listed in the Pfam database in January 2020) were retrieved. C-termini of effector domains and N-termini of NACHT/NB-ARC NLRs were extracted and – in both cases – only fragments between 10 and 150 aa were selected for further analysis. (This effectively excluded proteins with effector + NOD architectures.) The final set included around 200k of effector C-termini and 6.8k NLR N-termini (redundant).

**Identification of paired amyloid motifs.** Sets of N- and C-terminals were clustered using CD-HIT to reduce redundancy at the 70% similarity threshold (separately for each effector domain, together for NACHT and NB-ARC). Motif search was performed using MEME with the following parameters: -nmotifs 100, -minsites 1% of sequences but no less than 5 and no more than 10, -maxsites 500, -minw 10, -maxw 30, -mod anr. For each of the identified motifs, HMM profiles were built in a two-stage procedure as described above with an exception that initial profiles were used for searching against all N- and C-termini. The same-motif hits in effector proteins and NLRs were matched based on genomic proximity (up to 20 kbp) of genes encoding the proteins. Motifs with at least 3 non-redundant pairs of motif instances (Supplementary File 10) were then clustered on the basis of their co-occurrence in pairs of sequences.

Finally, ASM hits in N-termini of all NLRs investigated in this study and in C-termini of all effector domains found were matched on the strain level (through the BioSample and BioProject identifiers) in order to identify potentially correlated pairs, which are not co-localized in genomes (Supplementary File 11).

**Specialized searches for amyloid motifs in Basidiomycota.** MLKL\_NTD homologs were searched in UniProt [117] through the web-based hmmsearch [118] with standard parameters starting from the alignment of the largest BaMLKL cluster in Basidiomycota (representative protein: KIM77258), trimmed to the NACHT\_N match. Hits were further restricted to GenBank sequences with length up to 400 amino acids and no Pfam P-loop\_NTPase clan (CL0023) annotation at E-value of 1. C-termini (100aa) of resulting 242 BaMLKL homologs were scanned with the PCFG BASS model (Supplementary File 7) with the same parameters as above (except the minimum scanning window length of 15). For proteomes with the most promising hits in BaMLKL homologs (log10 score above 3, eight sequences from six species), N-termini (150 aa) of all NLR proteins were again scanned with the grammars. Promising N-terminal hits were obtained for *Moniliophthora roreri* (strains 2995 and 2997), *Laccaria amethystina* (strain LaAM-08-1), and *Fibularhizoctonia* sp. CBS 109695. The matched fragments were aligned with their C-terminal counterparts on the per genome basis with Mafft [125] in an accurate mode (`-maxiterate 1000 -localpair`). The NLR N-terminal and BaMLKL C-terminal motifs aligned satisfactorily for *M. roreri* (we only analyzed strain 2997 due to high similarity between the strains) and *Fibularhizoctonia* sp. CBS 109695. The alignments were then extended and trimmed manually (Fig. S4 and S5). In addition, the sequences were scanned with the 16 HMM profiles of amyloid-like motifs (E-value of  $1e-2$ ).

Eventually, fungal proteomes in UniProt were scanned using web-based jackhmmer [118] with standard parameters starting from the double HET-s motif from Q03689 (AAB94631) (residues 218–289) of *Podospora anserina*, which resulted in finding five complete HeLo-HRAM-HRAM proteins in two Agaricomycetes: *Sphaerobolus stellatus* SS14 and *Gymnopus luxurians* FD-317 M1. NLRs in these genomes were then scanned with the PCFG model and the hits exceeding the log10 score threshold of 2.33 were aligned with their C-terminal counterparts on the per genome basis with Mafft [125] in the accurate mode. Finally, the alignments were curated manually (poorly aligned sequences were excluded, sequences were extended or trimmed if necessary, Fig. S6 and S7).

**Visualization.** Basic data processing and visualization was conducted in Python using pandas [126, 127], matplotlib [128] and seaborn [129] packages, as well as in LibreOffice, GIMP and Inkscape. Multiple sequence alignments and logos were generated using TeXshade [130]. The graph of logos in Fig. S3 was generated with graphviz 2.40.1 [131]. Visualizations of structural models were generated with RasMol [132] (Fig. 2) or taken directly from the ColabFold notebook [70] (Fig. S2).

## Experimental methods

### In vitro analysis

**Peptide synthesis.** All commercially available reagents and solvents were purchased from Merck, Sigma-Aldrich and Lipopharm.pl, and used without further purification. Peptides EQB50682.1\_332\_355 (VFHGKGIQHTGSGNFSVGNLDSIS) and EQB50683.1\_9\_31 (FHGHGIALSGAGNITVGGDFIIG) were synthesized with an automated solid-phase peptide synthesizer (Liberty Blue, CEM) using rink amide AM resin (loading: 0.59 mmol/g). Fmoc deprotection was achieved using 20% piperidine in DMF for 1 min at 90°C. A double-coupling procedure was performed with 0.5 M solution of DIC and 0.25 M solution of OXYMA (1:1) in DMF for 4 min at 90°C. Cleavage of the peptides from the

resin was accomplished with the mixture of TFA/TIS/H<sub>2</sub>O (95:2.5:2.5) after 3 h of shaking. The crude peptide was precipitated with ice-cold Et<sub>2</sub>O and centrifuged (8000 rpm, 15 min, 2°C). Peptides were purified using preparative HPLC (Knauer Prep) with a C18 column (Thermo Scientific, Hypersil Gold 12  $\mu$ l, 250  $\times$  20 mm) with water/acetonitrile (0.05% TFA) eluent system.

**Peptide analytics.** Analytical high-performance liquid chromatography (HPLC) was performed using Kinetex 5 $\mu$  EVO C18 100A 150  $\times$  4.6 mm column. Program (eluent A: 0.05% TFA in H<sub>2</sub>O, eluent B: 0.05% TFA in acetonitrile, flow 0.5 mL/min): A: t=0 min, 90% A; t=45 min (25 min in case of EQB50682.1\_332\_355). Peptides were studied by WATERS LCT Premier XE System consisting of high resolution mass spectrometer (MS) with a time of flight (TOF).

**Attenuated Total Reflectance – Fourier Transform Infrared Spectroscopy (ATR-FTIR).** Lyophilized peptides were dissolved in D<sub>2</sub>O (deuterium oxide, 99.8% D, Carl Roth, GmbH, Germany) to a final concentration of 2 mg/ml. The spectroscopic measurements were performed directly after dissolving peptides in a solvent, after 7 and 40 days of incubation process at 37°C (98.6°F), and after 40 days of incubation at 4°C (39.2°F). Each time, 10  $\mu$ l of peptide solution was dropped directly on the diamond surface and was allowed to dry out. ATR-FTIR spectra were recorded using a Nicolet 6700 FTIR Spectrometer (Thermo Scientific, USA) with Golden Gate Mk II ATR Accessory with Heated Diamond Top-plate (PIKE Technologies). The spectrometer was continuously purged with dry air. Directly before sampling, the background spectrum of diamond/air was collected as a reference. For each spectrum 512 scans with a resolution of 4 cm<sup>-1</sup> were co-added. All spectra were obtained in the range of 4000–450 cm<sup>-1</sup> at 20°C (68.0°F).

**Spectroscopy data treatment.** ATR-FTIR spectra were initially preprocessed using OMNICTM software (version 8, Thermo Fisher Scientific, USA): atmospheric and ATR correction. All spectra were analyzed using the OriginPro (version 2019, OriginLab Corporation, USA). The analysis included: baseline correction, smoothing using the Savitzky-Golay polynomial filter (polynomial order 2, a window size of 9 points) [133] and normalization to 1 for the Amide II' band. Spectra in the amide bands region (1750–1500 cm<sup>-1</sup>) were deconvoluted into subcomponents using the Lorentz function based on second and fourth derivative spectra.

**Atomic Force Microscopy.** AFM images were acquired in tapping mode using a Nanoscope IIIa scanning probe microscope with Extender Module (Bruker) in the dynamic modus. An active vibration isolation platform was applied. Olympus etched silicon cantilevers were used with a typical resonance frequency in the range of 100–200 kHz and a spring constant of 40 N/m. The set-point amplitude of the cantilever was maintained by the feedback circuitry at 80% of the free oscillation amplitude of the cantilever. The samples with peptides were placed on freshly cleaved ultra-clean mica (Nano and More) and incubated at room temperature for 30 s. The mica discs were then rinsed with ultra-clean purified 18.2 M $\Omega$  deionized water and dried using gentle nitrogen gas flow. All samples were measured at room temperature in air. Structural analysis and height measurements of acquired images were performed with Nanoscope v.6.13 software.

**Thioflavin T (ThT) fluorescence assay.** ThT powder was dissolved in MilliQ to final concentration 2 mM and filtered through 0.22  $\mu$ m syringe. ThT solution was dissolved in 50 mM Tris-HCl

(pH= 7.4) to final concentration 10  $\mu$ M and filtered. The 90  $\mu$ L of ThT buffer was mixed with 10  $\mu$ L of peptide solution (concentration 400  $\mu$ M) in the 96-wells plate. Samples were measured on the SpectraMax® Gemini™ XPS Microplate (Molecular Devices LLC). The measurements were conducted in RT. The excitation wavelength was set at 450 nm and the emission was recorded in the range from 470 to 500 nm. Each group of experiment contained three parallel samples and the data were averaged after measurements.

## In vivo analysis

**Strains and plasmids.** The *Podospora anserina*  $\Delta$ hellp ( $\Delta$ Pa\_5\_8070)  $\Delta$ het-s ( $\Delta$ Pa\_3\_620)  $\Delta$ hellf ( $\Delta$ Pa\_3\_9900) strain [103] was used as recipient strain for the expression of molecular fusions of PUASM (PNP\_UDP-side C-terminal EQB50682.1\_332\_355 VFHGKGIQHTGSGNFSVGNLDSIS) from the plant pathogenic fungus *Colletotrichum gloeosporioides* Cg-14 [89] and the GFP (green fluorescent protein) or RFP (red fluorescent protein). These fusions were expressed from plasmids based on the pGEM-T backbone (Promega) named pOP [40] and containing either the GFP or RFP encoding gene, or in a derivative of the pAN52.1 GFP vector [134], named pGB6-GFP and containing the GFP encoding gene. In both cases, the molecular fusions were under the control of the strong constitutive *P. anserina* *gpd* (glyceraldehyde-3-phosphate dehydrogenase) promoter. The  $\Delta$ hellp  $\Delta$ het-s  $\Delta$ hellf strain was transformed as described [135] with a fusion construct along with a second vector carrying a ble phleomycin-resistance gene, pPaBle (using a 10:1 molar ratio). Phleomycin-resistant transformants were selected, grown for 30 h at 26°C and screened for the expression of the transgenes using fluorescence microscopy. PUASM was amplified with specific primers either 5' ggcttaatttaaATGGTCTTTCATGGCAAGGGCATCC 3' and 5' ggcagatcttgctccGGAGATGCTGAGATCG 3' for cloning in pOP plasmids, or 5' ggcgcggcggcGTCTTTCATGGCAAGGGCATC 3' and 5' ggcGGATCCTTAGGAGATGCTGAGATCGTTGCC 3' for cloning in the pGB6 plasmid (capital letters correspond to the PUASM sequence). The PCR products were cloned upstream of the GFP or RFP coding sequence in the pOP plasmids using PacI/BglII restriction enzymes to generate the pOPPUASM-GFP and pOPPUASM-RFP vectors in which in addition to the BglII site, a two amino acid linker (GA) was introduced between the sequences encoding PUASM and GFP or RFP and cloned downstream of the GFP using NotI/BamHI restriction enzymes to generate the pGB6-GFP-PUASM plasmid.

**Microscopy.** *P. anserina* hyphae were inoculated on solid medium and cultivated for 24 to 48 h at 26°C. The medium was then cut out, placed on a glass slide and examined with a Leica DMRXA microscope equipped with a Micromax CCD (Princeton Instruments) controlled by the Metamorph 5.06 software (Roper Scientific). The microscope was fitted with a Leica PL APO 63X immersion lens.

**Prion propagation.** Methods for determination of prion formation and propagation were previously described [136, 20]. Prion formation and propagation can be observed using microscopy by monitoring the formation of fluorescent dots. Spontaneous prion formation is first monitored as the rate of spontaneously acquired prion phenotype (dot formation) in the initially prion-free subculture after 5, 11, 18, 32, 49 and 75 days of growth at 26°C on corn-meal agar using microscopy as described. Prion formation can also be measured as the ability to propagate prions from a donor strain (containing prion) to a prion-free strain (induced strain). In practice, prion-free strains are confronted on solid corn-meal agar medium for 2 to 5 days (contact between strains were observed

after 24 to 36 hours of culture) before being subcultured and observed by fluorescence microscopy for the presence of dots, this test is referred as induced prion formation. At least 18 different transformants were used and the tests were realized in triplicates. It is to note that transformants were randomly tested for prion formation allowing various expression levels of the transgene (high levels of expression are usually associated with rapid spontaneous prion formation) except for the induced conversion test where transformants expressing moderate level of transgene were preferred to limit the rate of spontaneous transition within the timing of the experiment that could mask the prion induction.

**Author contributions.** **JWW:** data curation, formal analysis, investigation, methodology, software, writing; **ET:** data curation, formal analysis, investigation, methodology, software, validation, writing; **MG-G:** investigation, methodology, resources, validation, visualization, writing; **VC:** investigation, methodology, resources, validation, visualization; **NS:** investigation, validation, visualization, writing; **MS:** investigation, validation, resources, writing; **MK:** investigation, resources, supervision, validation, visualization, writing; **SJS:** conceptualization, funding acquisition, investigation, methodology, resources, validation, writing; **WD:** conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, supervision, validation, visualization, writing.

**Acknowledgments.** This research has been partially funded by National Science Centre, Poland, grant no. 2019/35/B/NZ2/03997 (PI: Prof. Małgorzata Kotulska) (WD, MG-G), by National Centre for Research and Development, Poland ([ncbr.gov.pl](http://ncbr.gov.pl)), project no. POWR.03.02.00-00-I003/16 (NS), and by Politechnika Wroclawska statutory funds, and supported by Wroclaw Centre for Networking and Supercomputing grant no. 98 and the E-SCIENCE.PL infrastructure. Work in the SJS lab was supported by an ANR grant (SFAS, R-17-CE11-0035).

## References

- [1] J. D. G. Jones, R. E. Vance, and J. L. Dangl, “Intracellular innate immune surveillance devices in plants and animals,” *Science*, vol. 354, no. 6316, p. aaf6395, 2016.
- [2] J. Uehling, A. Deveau, and M. Paoletti, “Do fungi have an innate immune response? an NLR-based comparison to plant and animal immune systems,” *PLoS Pathogens*, vol. 13, no. 10, p. e1006578, 2017.
- [3] Z. Duxbury, C.-h. Wu, and P. Ding, “A comparative overview of the intracellular guardians of plants and animals: Nlrs in innate immunity and beyond,” *Annual Review of Plant Biology*, vol. 72, no. 1, pp. 155–184, 2021.
- [4] E. Koonin and L. Aravind, “Origin and evolution of eukaryotic apoptosis: the bacterial connection,” *Cell Death & Differentiation*, vol. 9, pp. 394–404, 2002.
- [5] A. Daskalov, W. Dyrka, and S. J. Saupe, “NLR function in fungi as revealed by the study of self/non-self recognition systems,” in *Genetics and Biotechnology* (J. Benz, ed.), The Mycota, Springer, 3 ed., 2020.

- [6] Y. Xiong, Z. Han, and J. Chai, “Resistosome and inflammasome: platforms mediating innate immunity,” *Current opinion in plant biology*, vol. 56, pp. 47–55, August 2020.
- [7] G. Bi and J.-M. Zhou, “Regulation of cell death and signaling by pore-forming resistosomes,” *Annual review of phytopathology*, vol. 59, pp. 239–263, August 2021.
- [8] I. M. L. Saur, R. Panstruga, and P. Schulze-Lefert, “NOD-like receptor-mediated plant immunity: from structure to cell death,” *Nat. Rev. Immunol.*, vol. 21, pp. 305–318, May 2021.
- [9] E. A. van der Biezen and J. D. Jones, “The nb-arc domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals,” *Current Biology*, vol. 8, no. 7, pp. R226–R228, 1998.
- [10] E. Koonin and L. Aravind, “The nacht family — a new group of predicted ntpases implicated in apoptosis and mhc transcription activation.,” *Trends in Biochemical Sciences*, vol. 25, pp. 223–224, 2000.
- [11] S. Saupe, B. Turcq, and J. Begueret, “A gene responsible for vegetative incompatibility in the fungus *podospira anserina* encodes a protein with a gtp-binding motif and g beta homologous domain,” *Gene*, vol. 162, pp. 135–139, 1995.
- [12] M. Paoletti, “Vegetative incompatibility in fungi: From recognition to cell death, whatever does the trick,” *Fungal Biol. Rev.*, vol. 30, pp. 152–162, Oct. 2016.
- [13] A. P. Gonçalves, J. Heller, A. M. Rico-Ramírez, A. Daskalov, G. Rosenfield, and N. L. Glass, “Conflict, competition, and cooperation regulate social interactions in filamentous fungi,” *Annual Review of Microbiology*, vol. 74, no. 1, pp. 693–712, 2020.
- [14] W. Dyrka, M. Lamacchia, P. Durrens, B. Kobe, A. Daskalov, M. Paoletti, D. J. Sherman, and S. J. Saupe, “Diversity and variability of nod-like receptors in fungi,” *Genome Biology and Evolution*, vol. 6, pp. 3137–3158, 2014.
- [15] J. Urbach and F. Ausubel, “The NBS-LRR architectures of plant r-proteins and metazoan nlrs evolved in independent events,” *Proceedings of the National Academy of Sciences*, vol. 114, pp. 1063–1068, 2017.
- [16] E. Espagne, P. Balhadere, M. Penin, C. Barreau, and B. Turcq, “HET-E and HET-D belong to a new subfamily of WD40 proteins involved in vegetative incompatibility specificity in the fungus *podospira anserina*,” *Genetics*, vol. 161, pp. 71–81, 2002.
- [17] E. Bastiaans, A. Debets, D. Aanen, A. van Diepeningen, S. Saupe, and M. Paoletti, “Natural variation of heterokaryon incompatibility gene *het-c* in *podospira anserina* reveals diversifying selection,” *Molecular Biology and Evolution*, vol. 31, pp. 962–974, 2014.
- [18] J. Heller, C. Clavé, P. Gladieux, S. J. Saupe, and N. L. Glass, “NLR surveillance of essential SEC-9 SNARE proteins induces programmed cell death upon allorecognition in filamentous fungi,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, pp. E2292–E2301, Mar. 2018.
- [19] M. Paoletti and S. J. Saupe, “Fungal incompatibility: Evolutionary origin in pathogen defense?,” *BioEssays*, vol. 31, no. 11, pp. 1201–1210, 2009.

- [20] W. Dyrka, V. Coustou, A. Daskalov, A. Lends, T. Bardin, M. Berbon, B. Kauffmann, C. Blancard, B. Salin, A. Loquet, and S. J. Saupe, “Identification of nlr-associated amyloid signaling motifs in bacterial genomes,” *Journal of Molecular Biology*, 2020. in press.
- [21] M. Paoletti, S. J. Saupe, and C. Clavé, “Genesis of a fungal non-self recognition repertoire,” *PLoS ONE*, vol. 2, p. e283, 03 2007.
- [22] D. Chevanne, S. Saupe, C. Clave, and M. Paoletti, “Wd-repeat instability and diversification of the podospora anserina hnwD non-self recognition gene family,” *BMC Evolutionary Biology*, vol. 10, no. 1, p. 134, 2010.
- [23] O. J. Rando and K. J. Verstrepen, “Timescales of genetic and epigenetic inheritance,” *Cell*, vol. 128, pp. 655–668, 2007.
- [24] M. Iotti, A. Rubini, E. Tisserant, A. Kholer, F. Paolocci, and A. Zambonelli, “Self/nonsel recognition in tuber melanosporum is not mediated by a heterokaryon incompatibility system,” *Fungal Biology*, vol. 116, no. 2, pp. 261–275, 2012.
- [25] R. D. Finn, P. Coghill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate, and A. Bateman, “The pfam protein families database: towards a more sustainable future,” *Nucleic Acids Research*, 2016.
- [26] S. J. Saupe, “Amyloid signaling in filamentous fungi and bacteria,” *Annual Review of Microbiology*, vol. 74, no. 1, pp. 673–691, 2020.
- [27] A. Daskalov, B. Habenstein, D. Martinez, A. J. Debets, R. Sabate, A. Loquet, and S. J. Saupe, “Signal transduction by a fungal NOD-like receptor based on propagation of a prion amyloid fold,” *PLoS Biology*, vol. 13, no. 2, p. e1002059, 2015.
- [28] C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, and B. H. Meier, “Amyloid fibrils of the HET-s(218–289) prion form a beta solenoid with a triangular hydrophobic core,” *Science*, vol. 319, no. 5869, pp. 1523–1526, 2008.
- [29] J. Greenwald, C. Buhtz, C. Ritter, W. Kwiatkowski, S. Choe, M. L. Maddelein, F. Ness, S. Cescau, A. Soragni, D. Leitz, S. J. Saupe, and R. Riek, “The mechanism of prion inhibition by HET-S,” *Molecular Cell*, vol. 38, pp. 889–899, 2010.
- [30] H. van Melckebeke, C. Wasmer, A. Lange, E. AB, A. Loquet, A. Böckmann, and B. H. Meier, “Atomic-resolution three-dimensional structure of het-s(218–289) amyloid fibrils by solid-state nmr spectroscopy,” *Journal of the American Chemical Society*, vol. 132, no. 39, pp. 13765–13775, 2010.
- [31] C. Seuring, J. Greenwald, C. Wasmer, R. Wepf, S. J. Saupe, B. H. Meier, and R. Riek, “The mechanism of toxicity in HET-S/HET-s prion incompatibility,” *PLoS Biology*, vol. 10, no. 12, p. e1001451, 2012.
- [32] A. Daskalov, M. Paoletti, F. Ness, and S. J. Saupe, “Genomic clustering and homology between het-s and the nwd2 stand protein in various fungal genomes,” *PLoS ONE*, vol. 7, p. e34854, 04 2012.

- [33] A. Daskalov, W. Dyrka, and S. J. Saupe, “Theme and variations: evolutionary diversification of the HET-s functional amyloid motif,” *Scientific Reports*, vol. 5, p. 12494, July 2015.
- [34] A. Balguerie, S. Dos Reis, C. Ritter, S. Chaignepain, B. Couлары-Salin, V. Forge, K. Bathany, I. Lascu, J. M. Schmitter, R. Riek, and S. J. Saupe, “Domain organization and structure-function relationship of the het-s prion protein of *podospora anserina*,” *The EMBO Journal*, vol. 22, no. 9, pp. 2071–2081, 2003.
- [35] N. D. Fedorova, J. H. Badger, G. D. Robson, J. R. Wortman, and W. C. Nierman, “Comparative analysis of programmed cell death pathways in filamentous fungi,” *BMC Genomics*, vol. 6, p. 177, 2005.
- [36] A. Daskalov, D. Martinez, V. Coustou, N. El Mammeri, M. Berbon, L. B. Andreas, B. Bardiaux, J. Stanek, A. Noubhani, B. Kauffmann, J. S. Wall, G. Pintacuda, S. J. Saupe, B. Habenstein, and A. Loquet, “Structural and molecular basis of cross-seeding barriers in amyloids,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 1, 2021.
- [37] S. Graziani, P. Silar, and M. Daboussi, “Bistability and hysteresis of the ‘secteur’ differentiation are controlled by a two-gene locus in *nectria haematococca*,” *BMC Biology*, vol. 2, p. 18, 2004.
- [38] X. Sun, J. Yin, M. A. Starovasnik, W. J. Fairbrother, and V. M. Dixit, “Identification of a novel homotypic interaction motif required for the phosphorylation of receptor-interacting protein (rip) by rip3,” *The Journal of Biological Chemistry*, vol. 277, no. 11, pp. 9505–9511, 2002.
- [39] M. Rebsamen, L. X. Heinz, E. Meylan, M.-C. Michallet, K. Schroder, K. Hofmann, J. Vazquez, C. A. Benedict, and J. Tschopp, “Dai/zbp1 recruits rip1 and rip3 through rip homotypic interaction motifs to activate nf-kb,” *EMBO reports*, vol. 10, no. 8, pp. 916–922, 2009.
- [40] A. Daskalov, B. Habenstein, R. Sabaté, M. Berbon, D. Martinez, S. Chaignepain, B. Couлары-Salin, K. Hofmann, A. Loquet, and S. J. Saupe, “Identification of a novel cell death-inducing domain reveals that fungal amyloid-controlled programmed cell death is related to necroptosis,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 10, pp. 2720–2725, 2016.
- [41] A. V. Kajava, K. Klopffleisch, S. Chen, and K. Hofmann, “Evolutionary link between metazoan RHIM motif and prion-forming domain of fungal heterokaryon incompatibility factor HET-s/HET-s,” *Scientific Reports*, vol. 4, pp. 1–6, Dec. 2014.
- [42] A. B. Ahmed, N. Znassi, M.-T. Château, and A. V. Kajava, “A structure-based approach to predict predisposition to amyloidosis,” *Alzheimer’s & Dementia*, vol. 11, pp. 681–690, June 2015.
- [43] W. Dyrka, M. Gąsior-Głogowska, and M. Szeferczyk, “Searching for universal model of amyloid signaling motifs using probabilistic context-free grammars,” *bioRxiv*, p. 2021.01.22.426858, 2021.

- [44] J. Li, T. McQuade, A. B. Siemer, J. Napetschnig, K. Moriwaki, Y. S. Hsiao, E. Damko, D. Moquin, T. Walz, A. McDermott, F. K. Chan, and H. Wu, “The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis,” *Cell*, vol. 150, no. 2, pp. 339–350, 2012.
- [45] A. Kleino, N. F. Ramia, G. Bozkurt, Y. Shen, H. Nailwal, J. Huang, J. Napetschnig, M. Gangloff, F. K.-M. Chan, H. Wu, J. Li, and N. Silverman, “Peptidoglycan-sensing receptors trigger the formation of functional amyloids of the adaptor protein imd to initiate drosophila nf-kb signaling,” *Immunity*, vol. 47, no. 4, pp. 635 – 647.e6, 2017.
- [46] M. Steinegger and J. Soeding, “MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets,” *Nature Biotechnology*, vol. 35, pp. 1026–1028, 2017.
- [47] B. E. Suzek, Y. Wang, H. Huang, P. B. McGarvey, C. H. Wu, and the UniProt Consortium, “UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches,” *Bioinformatics*, vol. 31, pp. 926–932, 11 2014.
- [48] M. Mirdita, L. von den Driesch, C. Galiez, M. J. Martin, J. Söding, and M. Steinegger, “Uniclust databases of clustered and deeply annotated protein sequences and alignments,” *Nucleic Acids Research*, vol. 45, pp. D170–D176, 11 2016.
- [49] M. Remmert, A. Biegert, A. Hauser, and J. Soeding, “HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment,” *Nature Methods*, vol. 9, no. 2, pp. 173–175, 2012.
- [50] R. Oliva, J. Win, S. Raffaele, L. Boutemy, T. O. Bozkurt, A. Chaparro-Garcia, M. E. Segretin, R. Stam, S. Schornack, L. M. Cano, M. van Damme, E. Huitema, M. Thines, M. J. Banfield, and S. Kamoun, “Recent developments in effector biology of filamentous plant pathogens,” *Cellular Microbiology*, vol. 12, no. 6, pp. 705–715, 2010.
- [51] T. Liu, W. Ye, Y. Ru, X. Yang, B. Gu, K. Tao, S. Lu, S. Dong, X. Zheng, W. Shan, Y. Wang, and D. Dou, “Two host cytoplasmic effectors are required for pathogenesis of phytophthora sojae by suppression of host defenses,” *Plant physiology*, vol. 155, pp. 490–501, January 2011.
- [52] D. Zhang, A. M. Burroughs, N. D. Vidal, L. M. Iyer, and L. Aravind, “Transposons to toxins: the provenance, architecture and diversification of a widespread class of eukaryotic effectors,” *Nucleic Acids Research*, vol. 44, no. 8, pp. 3513–3533, 2016.
- [53] S. Voss, R. Betz, S. Heidt, N. Corradi, and N. Requena, “Ricr1, a crinkler effector from the arbuscular mycorrhizal fungus rhizophagus irregularis, functions in arbuscule development,” *Frontiers in microbiology*, vol. 9, p. 2068, 2018.
- [54] A. Hetmann and S. Kowalczyk, “Supresja odporności podstawowej typu PTI przez syntetyzowane w fitopatogenach białka efektorowe wprowadzane do wnętrza komórek infekowanej rośliny,” *Postępy Biochemii*, vol. 65, no. 1, pp. 58–71, 2019.

- [55] S. J. Grimshaw, H. R. Mott, K. M. Stott, P. R. Nielsen, K. A. Evetts, L. J. Hopkins, D. Nietlispach, and D. Owen, “Structure of the sterile alpha motif (sam) domain of the *saccharomyces cerevisiae* mitogen-activated protein kinase pathway-modulating protein ste50 and analysis of its interaction with the ste11 sam,” *The Journal of Biological Chemistry*, vol. 279, no. 3, pp. 2192–2201, 2004.
- [56] S. Ray, L. Chee, D. R. Matson, N. Y. Palermo, E. H. Bresnick, and K. J. Hewitt, “Sterile  $\alpha$ -motif domain requirement for cellular signaling and survival,” *Journal of Biological Chemistry*, vol. 295, no. 20, pp. 7113–7125, 2020.
- [57] H. Xiao, M. Kalman, K. Ikehara, S. Zemel, G. Glaser, and M. Cashel, “Residual guanosine 3',5'-bispyrophosphate synthetic activity of relA null mutants can be eliminated by spoT null mutations.,” *The Journal of Biological Chemistry*, vol. 266, no. 9, pp. 5980–5990, 1991.
- [58] S. K. Hanks and A. M. Quinn, “Protein kinase catalytic domain sequence database: Identification of conserved features of primary structure and classification of family members,” in *Protein Phosphorylation Part A: Protein Kinases: Assays, Purification, Antibodies, Functional Analysis, Cloning, and Expression*, vol. 200 of *Methods in Enzymology*, pp. 38–62, Academic Press, 1991.
- [59] L. Aravind and E. V. Koonin, “Classification of the caspase–hemoglobinase fold: Detection of new families and implications for the origin of the eukaryotic separins,” *Proteins: Structure, Function, and Bioinformatics*, vol. 46, no. 4, pp. 355–367, 2002.
- [60] T. P. Bonnert, K. E. Garka, P. Parnet, G. Sonoda, J. R. Testa, and J. E. Sims, “The cloning and characterization of human myd88: a member of an il-1 receptor related family 1,” *FEBS Letters*, vol. 402, no. 1, pp. 81–84, 1997.
- [61] M. A. Armant and M. J. Fenton, “Toll-like receptors: a family of pattern-recognition receptors in mammals,” *Genome Biology*, vol. 3, no. 8, p. REVIEWS3011, 2002.
- [62] C. Ponting and P. Parker, “Extending the c2 domain family: C2s in pkcs delta, epsilon, eta, theta, phospholipases, gaps, and perforin,” *Protein Science*, vol. 5, no. 1, pp. 162–166, 1996.
- [63] D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag, J. L. Sussman, K. H. Verschuere, and A. Goldman, “The alpha/beta hydrolase fold,” *Protein Engineering, Design and Selection*, vol. 5, pp. 197–211, 04 1992.
- [64] S. Detke, “Cloning of the *candida albicans* nucleoside transporter by complementation of nucleoside transport-deficient *saccharomyces*,” *Yeast*, vol. 14, no. 14, pp. 1257–1265, 1998.
- [65] S. Lu, J. Wang, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, M. Yang, D. Zhang, C. Zheng, C. J. Lanczycki, and A. Marchler-Bauer, “CDD/SPARCLE: the conserved domain database in 2020,” *Nucleic Acids Research*, vol. 48, pp. D265–D268, 11 2019.
- [66] J. Soeding, “Protein homology detection by HMM-HMM comparison,” *Bioinformatics*, vol. 21, no. 7, pp. 951–960, 2005.

- [67] L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, and V. Alva, “A completely reimplemented mpi bioinformatics toolkit with a new hhpred server at its core,” *Journal of Molecular Biology*, vol. 430, no. 15, pp. 2237–2243, 2018. Computation Resources for Molecular Biology.
- [68] M. Rübhelke, D. Fiegen, M. Bauer, F. Binder, J. Hamilton, J. King, S. Thamm, H. Nar, and M. Zeeb, “Locking mixed-lineage kinase domain-like protein in its auto-inhibited state prevents necroptosis,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 52, pp. 33272–33281, 2020.
- [69] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, “Highly accurate protein structure prediction with alphafold,” *Nature*, vol. 596, pp. 583–589, August 2021.
- [70] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, “Colabfold - making protein folding accessible to all,” *bioRxiv*, 2021.
- [71] Y. Zhang and J. Skolnick, “TM-align: a protein structure alignment algorithm based on the TM-score,” *Nucleic Acids Research*, vol. 33, pp. 2302–2309, 01 2005.
- [72] H. Adachi, M. P. Contreras, A. Harant, C.-h. Wu, L. Derevnina, T. Sakai, C. Duggan, E. Moratto, T. O. Bozkurt, A. Maqbool, J. Win, and S. Kamoun, “An n-terminal motif in nlr immune receptors is functionally conserved across distantly related plant species,” *eLife*, vol. 8, p. e49956, nov 2019.
- [73] D. E. McNamara, C. M. Dovey, A. T. Hale, G. Quarato, C. R. Grace, C. D. Guibao, J. Diep, A. Nourse, C. R. Cai, H. Wu, R. C. Kalathur, D. R. Green, J. D. York, J. E. Carette, and T. Moldoveanu, “Direct activation of human mlkl by a select repertoire of inositol phosphate metabolites,” *Cell Chemical Biology*, vol. 26, no. 6, pp. 863–877.e7, 2019.
- [74] S. Ghabrial and M. Nibert, “Victorivirus, a new genus of fungal viruses in the family totiviridae,” *Arch Virol*, no. 154, pp. 373–379, 2009.
- [75] A. M. Jethon, *Potential virual symbiont in leaf cutter ant system*. PhD thesis, The Pennsylvania State University, 2018.
- [76] Y. Jo, H. Choi, H. Chu, and W. K. Cho, “Identification of viruses from fungal transcriptomes,” *bioRxiv*, 2020.
- [77] G. Polekhina, A. Gupta, B. J. W. van Denderen, S. C. Feil, B. E. Kemp, D. Stapleton, and M. W. Parker, “Structural basis for glycogen recognition by amp-activated protein kinase,” *Structure (London, England : 1993)*, vol. 13, pp. 1453–1462, October 2005.

- [78] M. Novatchkova, A. Leibbrandt, J. Werzowa, A. Neubüser, and F. Eisenhaber, “The stir-domain superfamily in signal transduction, development and immunity,” *Trends in biochemical sciences*, vol. 28, pp. 226–229, May 2003.
- [79] B. Wu, J. Gong, L. Liu, T. Li, T. Wei, and Z. Bai, “Evolution of prokaryotic homologues of the eukaryotic sefir protein domain,” *Gene*, vol. 492, pp. 160–166, January 2012.
- [80] F. Jacob, S. Vernaldi, and T. Maekawa, “Evolution and conservation of plant nlr functions,” *Frontiers in Immunology*, vol. 4, p. 297, 2013.
- [81] Y. Guo and K. Narisawa, “Fungus-bacterium symbionts promote plant health and performance,” *Microbes and Environments*, vol. 33, no. 3, pp. 239–241, 2018.
- [82] T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, and W. S. Noble, “MEME Suite: tools for motif discovery and searching,” *Nucleic Acids Research*, vol. 37, no. suppl\_2, pp. W202–W208, 2009.
- [83] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. T. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, “The Protein Data Bank,” *Nucleic Acid Research*, vol. 28, pp. 235–242, 2000.
- [84] V. Krishnan, Y. Xu, K. Macon, J. E. Volanakis, and S. V. L. Narayana, “The structure of C2b, a fragment of complement component C2 produced during C3 convertase formation,” *Acta Crystallographica Section D*, vol. 65, pp. 266–274, Mar 2009.
- [85] X.-H. Bai, H.-J. Chen, Y.-L. Jiang, Z. Wen, Y. Huang, W. Cheng, Q. Li, L. Qi, J.-R. Zhang, Y. Chen, and C.-Z. Zhou, “Structure of pneumococcal peptidoglycan hydrolase lytB reveals insights into the bacterial cell wall remodeling and pathogenesis,” *Journal of Biological Chemistry*, vol. 289, no. 34, pp. 23403–23416, 2014.
- [86] S. R. Eddy, “Accelerated profile HMM searches,” *PLoS Computational Biology*, vol. 7, p. e1002195, 10 2011.
- [87] W. Li and A. Godzik, “Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences,” *Bioinformatics*, vol. 22, pp. 1658–1659, 2006.
- [88] L. Fu, B. Niu, Z. Zhu, S. Wu, and W. Li, “CD-HIT: accelerated for clustering the next-generation sequencing data,” *Bioinformatics*, vol. 28, no. 23, pp. 3150–3152, 2012.
- [89] N. Alkan, X. Meng, G. Friedlander, E. Reuveni, S. Sukno, A. Sherman, M. Thon, R. Fluhr, and D. Prusky, “Global aspects of pacc regulation of pathogenicity genes in colletotrichum gloeosporioides as revealed by transcriptome analysis,” *Molecular Plant-Microbe Interactions*, vol. 26, no. 11, pp. 1345–1358, 2013.
- [90] P. M. Martins, S. Navarro, A. Silva, M. F. Pinto, Z. Sárkány, F. Figueiredo, P. J. B. Pereira, F. Pinheiro, Z. Bednarikova, M. Burdukiewicz, O. V. Galzitskaya, Z. Gazova, C. M. Gomes, A. Pastore, L. C. Serpell, R. Skrabana, V. Smirnovas, M. Ziaunys, D. E. Otzen, S. Ventura, and S. Macedo-Ribeiro, “Mirragge — minimum information required for reproducible aggregation experiments,” *Frontiers in Molecular Neuroscience*, vol. 13, p. 222, 2020.

- [91] R. Sarroukh, E. Goormaghtigh, J.-M. Ruyschaert, and V. Raussens, “Atr-ftir: a "rejuvenated" tool to investigate amyloid proteins,” *Biochimica et biophysica acta*, vol. 1828, p. 2328—2338, October 2013.
- [92] B. Shivu, S. Seshadri, J. Li, K. A. Oberg, V. N. Uversky, and A. L. Fink, “Distinct  $\beta$ -sheet structure in protein aggregates determined by atr-ftir spectroscopy,” *Biochemistry*, vol. 52, no. 31, pp. 5176–5183, 2013.
- [93] J. Ruyschaert and V. Raussens, “ATR-FTIR analysis of amyloid proteins,” *Methods Mol. Biol.*, vol. 1777, pp. 69–81, 2018.
- [94] F. S. Ruggeri, T. Šneideris, M. Vendruscolo, and T. P. Knowles, “Atomic force microscopy for single molecule characterisation of protein aggregation,” *Archives of Biochemistry and Biophysics*, vol. 664, pp. 134–148, 2019.
- [95] M. Biancalana and S. Koide, “Molecular mechanism of thioflavin-t binding to amyloid fibrils,” *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1804, no. 7, pp. 1405–1412, 2010.
- [96] C. Xue, T. Y. Lin, D. Chang, and Z. Guo, “Thioflavin t as an amyloid dye: fibril quantification, optimal concentration and effect on aggregation,” *Royal Society Open Science*, vol. 4, no. 1, p. 160696, 2017.
- [97] M. R. Nilsson, “Techniques to study amyloid fibril formation in vitro,” *Methods*, vol. 34, no. 1, pp. 151–160, 2004. *Investigating Protein Folding, Misfolding and Nonnative States: Experimental and Theoretical Methods*.
- [98] H. Li, F. Rahimi, S. Sinha, P. Maiti, G. Bitan, and K. Murakami, *Amyloids and Protein Aggregation—Analytical Methods*, ch. —, pp. —. American Cancer Society, 2009.
- [99] N. Wilkosz, M. Czaja, S. Seweryn, K. Skirlińska-Nosek, M. Szymonski, E. Lipiec, and K. Sofińska, “Molecular spectroscopic markers of abnormal protein aggregation,” *Molecules*, vol. 25, no. 11, 2020.
- [100] K. Berthelot, H. P. Ta, J. Géan, S. Lecomte, and C. Cullin, “In vivo and in vitro analyses of toxic mutants of HET-s: FTIR antiparallel signature correlates with amyloid toxicity,” *Journal of Molecular Biology*, vol. 412, no. 1, pp. 137–152, 2011.
- [101] J. R. Requena and H. Wille, “The structure of the infectious prion protein,” *Prion*, vol. 8, no. 1, pp. 60–66, 2014.
- [102] S. J. Roeters, A. Iyer, G. Pletikapić, V. Kogan, V. Subramaniam, and S. Woutersen, “Evidence for intramolecular antiparallel beta-sheet structure in alpha-synuclein fibrils from a combination of two-dimensional infrared spectroscopy and atomic force microscopy,” *Scientific Reports*, vol. 7, p. 41051, 2017.
- [103] T. Bardin, A. Daskalov, S. Barrouilhet, A. Granger-Farbos, B. Salin, C. Blancard, B. Kauffmann, S. J. Saupe, V. Coustou, and R. B. Wickner, “Partial prion cross-seeding between fungal and mammalian amyloid signaling motifs,” *mBio*, vol. 12, no. 1, pp. e02782–20, 2021.

- [104] W. Dyrka and J.-C. Nebel, “A stochastic context free grammar based framework for analysis of protein sequences,” *BMC Bioinformatics*, vol. 10, p. 323, 2009.
- [105] G. W. Wilburn and S. R. Eddy, “Remote homology search with hidden Potts models,” *PLOS Computational Biology*, vol. 16, no. 11, pp. 1–22, 2020.
- [106] A. P. Muntoni, A. Pagnani, M. Weigt, and F. Zamponi, “Aligning biological sequences by exploiting residue conservation and coevolution,” *Phys. Rev. E*, vol. 102, p. 062409, Dec 2020.
- [107] H. Talibart and F. Coste, “Ppalgn: optimal alignment of potts models representing proteins with direct coupling information,” *BMC Bioinformatics*, vol. 22, p. 317, June 2021.
- [108] W. Merrill, G. Weiss, Y. Goldberg, R. Schwartz, N. A. Smith, and E. Yahav, “A formal hierarchy of RNN architectures,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (Online), pp. 443–459, Association for Computational Linguistics, July 2020.
- [109] S. Bhattamishra, A. Patel, and N. Goyal, “On the computational power of transformers and its implications in sequence modeling,” in *Proceedings of the 24th Conference on Computational Natural Language Learning*, (Online), pp. 455–475, Association for Computational Linguistics, Nov. 2020.
- [110] W. Merrill, “Formal language theory meets modern NLP,” *CoRR*, vol. abs/2102.10094, 2021.
- [111] A. Nambiar, M. Heflin, S. Liu, S. Maslov, M. Hopkins, and A. Ritz, “Transforming the language of life: Transformer neural networks for protein prediction tasks,” in *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB ’20, (New York, NY, USA), Association for Computing Machinery, 2020.
- [112] N. Brandes, D. Ofer, Y. Peleg, N. Rappoport, and M. Linial, “Proteinbert: A universal deep-learning model of protein sequence and function,” *bioRxiv*, 2021.
- [113] M. Paoletti and C. Clavé, “The fungus-specific HET domain mediates programmed cell death in *Podospora anserina*,” *Eukaryotic Cell*, vol. 6, no. 11, pp. 2001–2008, 2007.
- [114] N. Szulc, M. Gąsior-Głogowska, J. W. Wojciechowski, M. Szefczyk, A. M. Żak, M. Burdukiewicz, and M. Kotulska, “Variability of amyloid propensity in imperfect repeats of csga protein of salmonella enterica and escherichia coli,” *International Journal of Molecular Sciences*, vol. 22, no. 10, 2021.
- [115] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, “Gapped blast and psi-blast: a new generation of protein database search programs,” *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–402, 1997.
- [116] F. Sievers, A. Wilm, D. G. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins, “Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega,” *Molecular Systems Biology*, vol. 7, p. 539, 2011.

- [117] T. U. Consortium, “UniProt: the universal protein knowledgebase in 2021,” *Nucleic Acids Research*, vol. 49, pp. D480–D489, 11 2020.
- [118] S. C. Potter, A. Luciani, S. R. Eddy, Y. Park, R. Lopez, and R. D. Finn, “HMMER web server: 2018 update,” *Nucleic Acids Research*, vol. 46, pp. W200–W204, 06 2018.
- [119] W. Dyrka, M. Pyzik, F. Coste, and H. Talibert, “Estimating probabilistic context-free grammars for proteins using contact map constraints,” *PeerJ*, vol. 7, p. e6559, Mar. 2019.
- [120] T. L. Bailey and C. Elkan, “Fitting a mixture model by expectation maximization to discover motifs in biopolymers,” in *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28–36, AAAI Press, Menlo Park, California, 1994.
- [121] N. R. Coordinators, “Database resources of the National Center for Biotechnology Information,” *Nucleic Acids Research*, vol. 46, pp. D8–D13, 11 2017.
- [122] P. W. Harrison, A. Ahamed, R. Aslam, B. T. F. Alako, J. Burgin, N. Buso, M. Courtot, J. Fan, D. Gupta, M. Haseeb, S. Holt, T. Ibrahim, E. Ivanov, S. Jayathilaka, V. Balavenkataraman Kadhivelu, M. Kumar, R. Lopez, S. Kay, R. Leinonen, X. Liu, C. O’Cathail, A. Pakseresht, Y. Park, S. Pesant, N. Rahman, J. Rajan, A. Sokolov, S. Vijayaraja, Z. Waheed, A. Zyoud, T. Burdett, and G. Cochrane, “The European Nucleotide Archive in 2020,” *Nucleic Acids Research*, vol. 49, pp. D82–D85, 11 2020.
- [123] K. Reitz, “Requests: HTTP for humans.” [www.python-requests.org](http://www.python-requests.org), 2017. [Online; accessed 2017-02-20].
- [124] M. Blech, “martinblech/xmltodict: Python module that makes working with XML feel like you are working with JSON.” World Wide Web electronic publication, 2005. <https://github.com/martinblech/xmltodict>.
- [125] K. Katoh and D. M. Standley, “MAFFT multiple sequence alignment software version 7: improvements in performance and usability,” *Molecular Biology and Evolution*, vol. 30, pp. 772–80, 2013.
- [126] Wes McKinney, “Data Structures for Statistical Computing in Python,” in *Proceedings of the 9th Python in Science Conference* (Stéfan van der Walt and Jarrod Millman, eds.), pp. 56–61, 2010.
- [127] T. pandas development team, “pandas-dev/pandas: Pandas,” Feb. 2020.
- [128] J. D. Hunter, “Matplotlib: A 2d graphics environment,” *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
- [129] M. L. Waskom, “seaborn: statistical data visualization,” *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021.
- [130] E. Beitz, “TeXshade: shading and labeling of multiple sequence alignments using LaTeX2e,” *Bioinformatics*, vol. 16, pp. 135–139, 02 2000.

- [131] E. R. Gansner and S. C. North, “An open graph visualization system and its applications to software engineering,” *Software Practice and Experience*, vol. 30, no. 11, pp. 1203–33, 2000.
- [132] R. A. Sayle and E. Milner-White, “Rasmol: biomolecular graphics for all,” *Trends in Biochemical Sciences*, vol. 20, no. 9, pp. 374–376, 1995.
- [133] A. Savitzky and M. Golay, “Smoothing and differentiation of data by simplified least squares procedures,” *Analytical Chemistry*, vol. 36, pp. 1627–1639, 1964.
- [134] A. Balguerie, S. Dos Reis, B. Couly-Salin, S. Chaignepain, M. Sabourin, J.-M. Schmitter, and S. J. Saupe, “The sequences appended to the amyloid core region of the HET-s prion protein determine higher-order aggregate organization in vivo,” *Journal of Cell Science*, vol. 117, pp. 2599–2610, 05 2004.
- [135] T. Bergès and C. Barreau, “Heat shock at an elevated temperature improves transformation efficiency of protoplasts from *podospora anserina*,” *Microbiology*, vol. 135, no. 3, pp. 601–604, 1989.
- [136] L. Benkemoun, R. Sabaté, L. Malato, S. D. Reis, H. Dalstra, S. J. Saupe, and M.-L. Maddelein, “Methods for the in vivo and in vitro analysis of [het-s] prion infectivity,” *Methods*, vol. 39, no. 1, pp. 61–67, 2006. Identification, analysis and characterization of fungal prions.

## Tables

Table 1: **N-terminal annotations of fungal NLRs.** Annotations based on Pfam [25] and inhouse [14] families and superfamilies. HHblits hits based on Pfam only. Direct hits include also inhouse annotations from [5]. Annotations are generally sorted by decreasing number of sequences. Homologous families are grouped together. Given are also the typical architecture of central and C-terminal domains if there is one clearly dominating, and the major taxonomic branch including vast majority of the hits.

Annotation	HHblits	Direct	Typical arch.	Major tax. branch
Goodbye-like	4912	4876	NACHT var.	Asco- & Basidiomycota
C2 Goodbye-like	356	55	NACHT WD	Agaricomycetes
PUP Goodbye-like	22	25	NACHT undef.	Eurotiomycetes
SesB-like Goodbye-like	20	18	NACHT ZF	Ascomycota
Goodbye-like/HeLo-like	130	—	NACHT undef.	Basidiomycota
HeLo-like	4044	4395	NACHT var.	Asco- & Basidiomycota
HeLo/HeLo-like	251	—	NACHT ANK	Ascomycota
HeLo	92	296	NACHT undef.	Ascomycota
SesB-like	3000	3851	var.	Ascomycota
PUP	3367	2647	var.	Ascomycota
C-terminus of PUP	45	—	NACHT var.	Ascomycota
HET	873	887	var.	Ascomycota
TIR	—	19	NACHT var.	Chytridiomycota
Patatin	646	651	NB-ARC TPR	Asco- & Basidiomycota
PFD-like	94	167	var.	Ascomycota
C2	—	137	NACHT WD	Agaricomycetes
RelA_SpoT	—	67	NACHT WD	Sordariomycetes
Crinkler	24	27	undef.	Glomeromycetes
SAM_Ste50p	24	—	undef.	Eurotiomycetes
Peptidase_S8	22	22	NACHT WD	Ascomycota
PKinase	—	22	NACHT var.	Ascomycota
CHAT	—	19	var.	Ascomycota

**Table 2: Unannotated N-terminal domains of fungal NLRs.** For each cluster given are: an accession of its representative protein is given (Rep. acc.), proposed annotation label, number of sequences in cluster in total (#seq) and non-redundant at identity threshold of 70% (nr70), median sequence length (Len.), typical architecture of central and C-terminal domains (undef.—unannotated, var.—variable), and taxonomic distribution. NLR\_PRDR stands for NLR effector domain with a Proline-Rich Disordered Region. See main text for details and Fig. S2 for structural models of KEY84097, KFH66451 and PQE30996.

Rep. acc.	Prop. annot.	#seq (nr70)	Len.	Typical arch.	Tax. distrib.
KEY84097	NLR_Helical	67 (42)	515	NACHT TPR	Pezizomycotina
KFH66451	NLR_Helical	27 (19)	617	NACHT WD	Mortierellomycetes
XP_007333708	NLR_PRDR	41 (32)	246	NACHT undef.	Agaricaceae
KIL58680	NLR_Koide	31 (15)	210	NACHT WD	<i>Amanita muscaria</i>
PQE30996	TIR-like	29 (19)	389	NACHT var.	Pezizomycotina

**Table 3: Amyloid-like motifs in short N-termini of NLRs.** Motif id indicates ranks in the MEME output. Motifs are grouped based on overlapping hits in NLRs and similar sequence patterns. Established and proposed motif annotation labels are given where applicable. #NLR and #nei. indicate number of sequences with a given motif in short N-termini of NLRs and C-termini of their genomic neighbors, respectively. AC+ indicates a proportion of motif instances for which ArchCandy score is 0.56 or above. Major taxonomic branch including vast majority of NLRs with the motif is given. #eff. indicates total number of single domain effector proteins (with established association to NLRs) with a given motifs in C-termini [5]. #cooc. indicates number of sequences with a given motif in short N-termini of NLRs / C-termini of effector proteins cooccurring in the same strains (genome assemblies). #str. is a number of such strains with cooccurrence.

Mot. id	Annot.	Len.	#NLR	AC+	Major tax.	#nei.	#eff.	#cooc.	#str.
NLR05	—	17	387	0.28	Agaricomycetes	0	0	—	—
NLR08	—	30	138	0.55	Agaricomycetes	0	0	—	—
NLR22	—	24	263	0.53	Agaricomycetes	0	0	—	—
NLR29	—	20	60	0.53	Agaricales	0	0	—	—
NLR44	—	12	24	0.00	Agaricomycetes	0	0	—	—
NLR07	PP	23	296	0.64	Ascomycota	37	112	128/58	45
NLR39	PP	24	20	0.80	Ascomycota	0	1	0/0	0
NLR12	HRAM3	26	110	0.48	Sordariomycetes	4	24	80/14	13
NLR13	HRAM	24	131	0.57	Ascomycota	16	50	37/21	18
NLR40	HRAM	26	24	0.92	Ascomycota	0	0	—	—
NLR17	—	26	24	0.88	<i>A. muscaria</i>	0	0	—	—
NLR19	—	27	53	0.60	<i>Tuber</i>	0	0	—	—
NLR20	—	21	37	0.68	Agaricales	0	0	—	—
NLR28	$\sigma$	28	71	0.80	Ascomycota	42	90	42/40	37
NLR32	PUASM	22	33	0.45	Sordariomycetes	11	55	25/21	18
NLR34	—	28	29	0.52	Ascomycota	0	0	—	—



## Figures

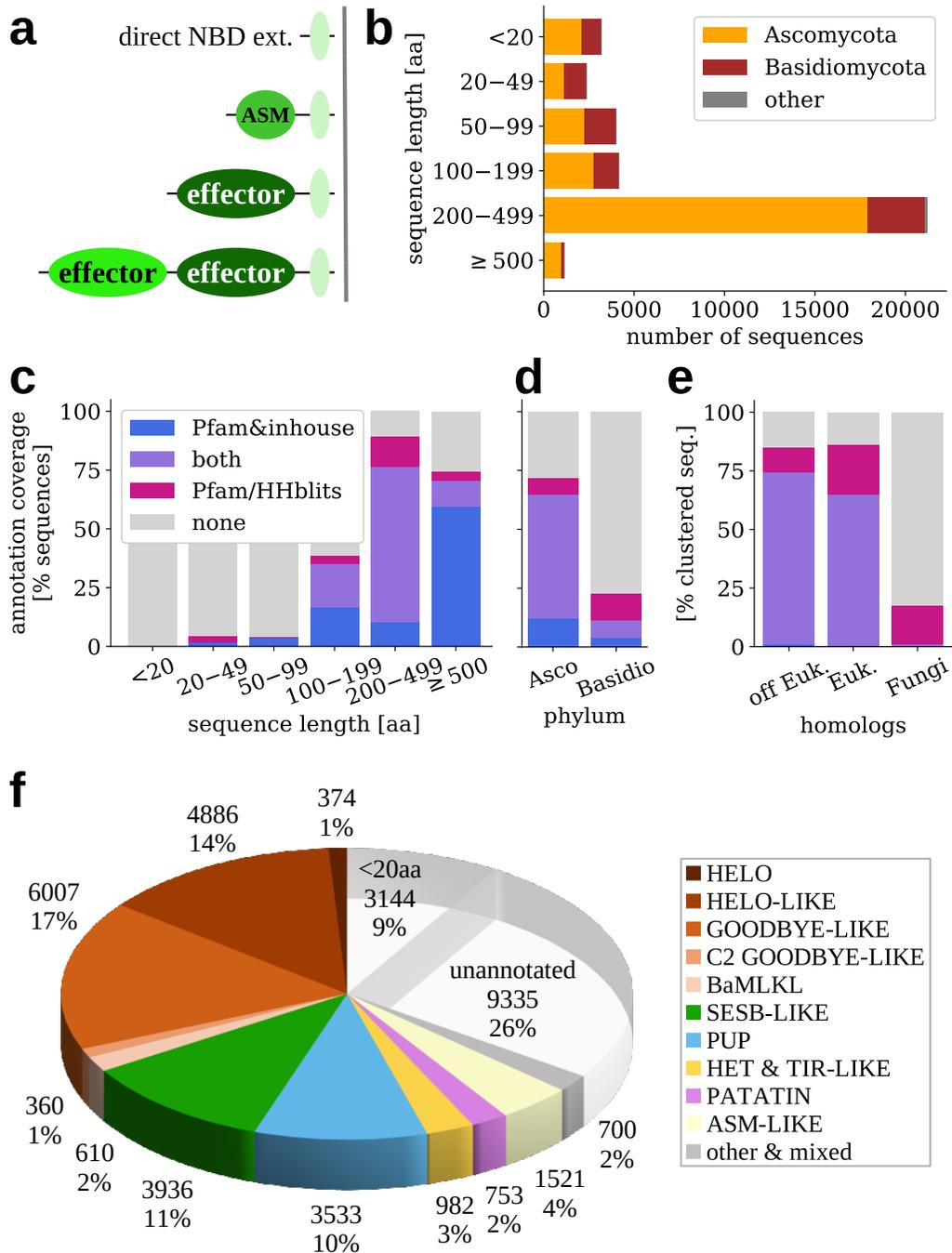


Figure 1: **Fungal NLR N-termini.** **a)** Major general architectures; **b)** Sequence length distribution; Annotation coverage with regard to **c)** sequence length, **d)** taxonomic division; **e)** Annotation coverage of clustered sequences (the 127 clusters, see main text) with regard to presence of taxonomically distant homologs in UniRef100 top hits. Euk. denotes *Eukaryota*. Colored bars indicate fraction of Pfam & inhouse annotated sequences (blue: only direct Pfam hits, violet: direct and with clustering & HHblits, rose: only with clustering & HHblits). Inhouse profiles were used only for direct Pfam searches. **f)** Distribution of domain families. Additional non-Pfam annotations included, see Results and Methods.

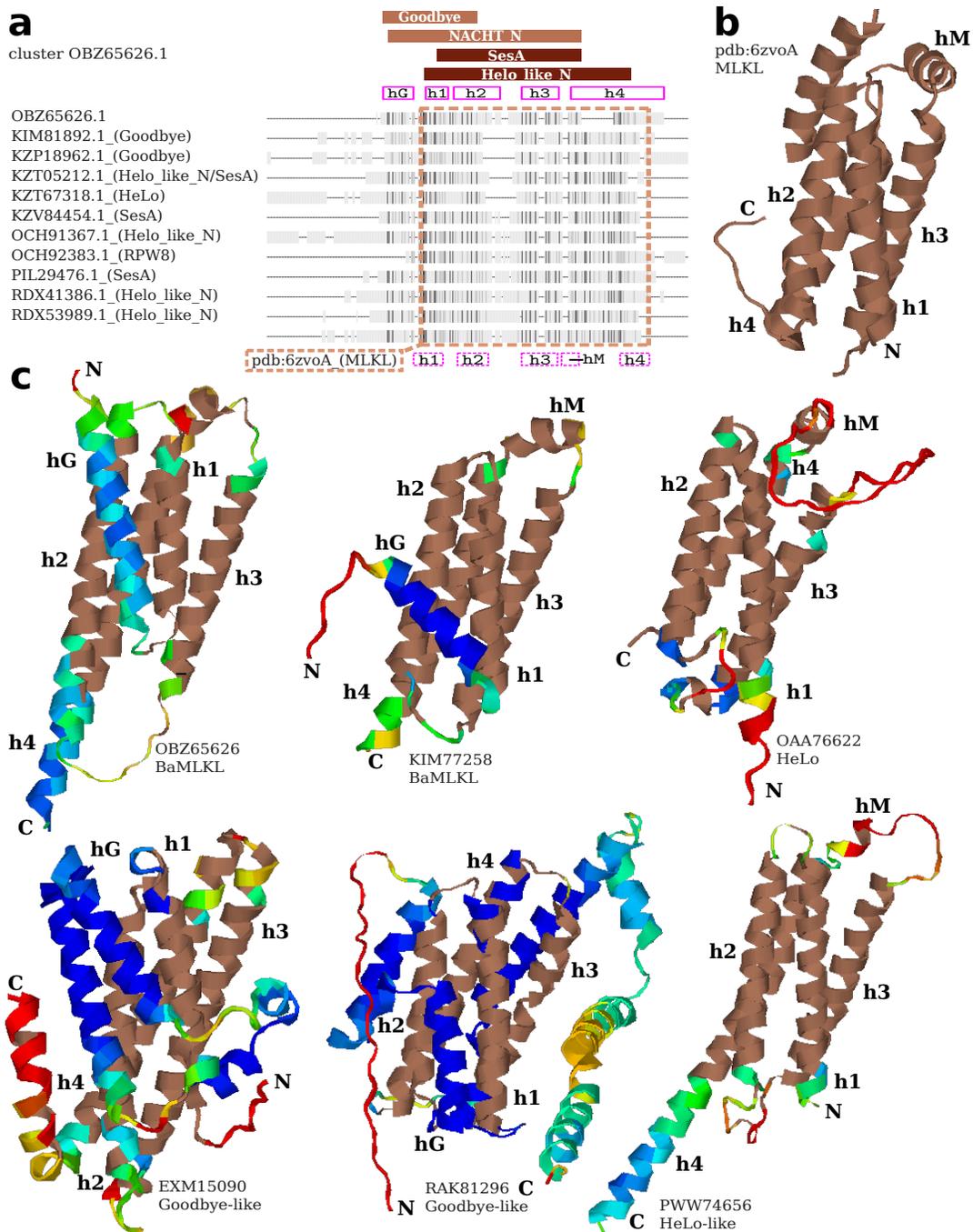


Figure 2: **MLKL-like N-termini.** (a) Fingerprint alignment of the doubly (Goodbye-like & Helo-like) annotated OBZ65626 cluster including non-redundant sequences with direct Pfam annotations. The alignment was truncated C-terminally. Darker shade implies higher conservation, while gaps are represented as lines. Columns matched with Pfam profiles of MLKL-like domains are indicated with brown bars. Columns corresponding to helices in a predicted OBZ65626 model are indicated with solid magenta boxes. Columns alignable to the human MLKL structure are framed with a brown dashed line. Columns corresponding to helices in the aligned MLKL structure are indicated with dashed magenta boxes. (b) The human MLKL structure (pdb:6zvoA). (c) Structural models of various MLKL-like domains predicted with AlphaFold2. Regions aligned to the human MLKL structure with TMalign are shown in brown. Rainbow colors indicate model quality in terms of IDDT (below or 50: red, 60: yellow, 70: green, 80: cyan, above 90: blue).

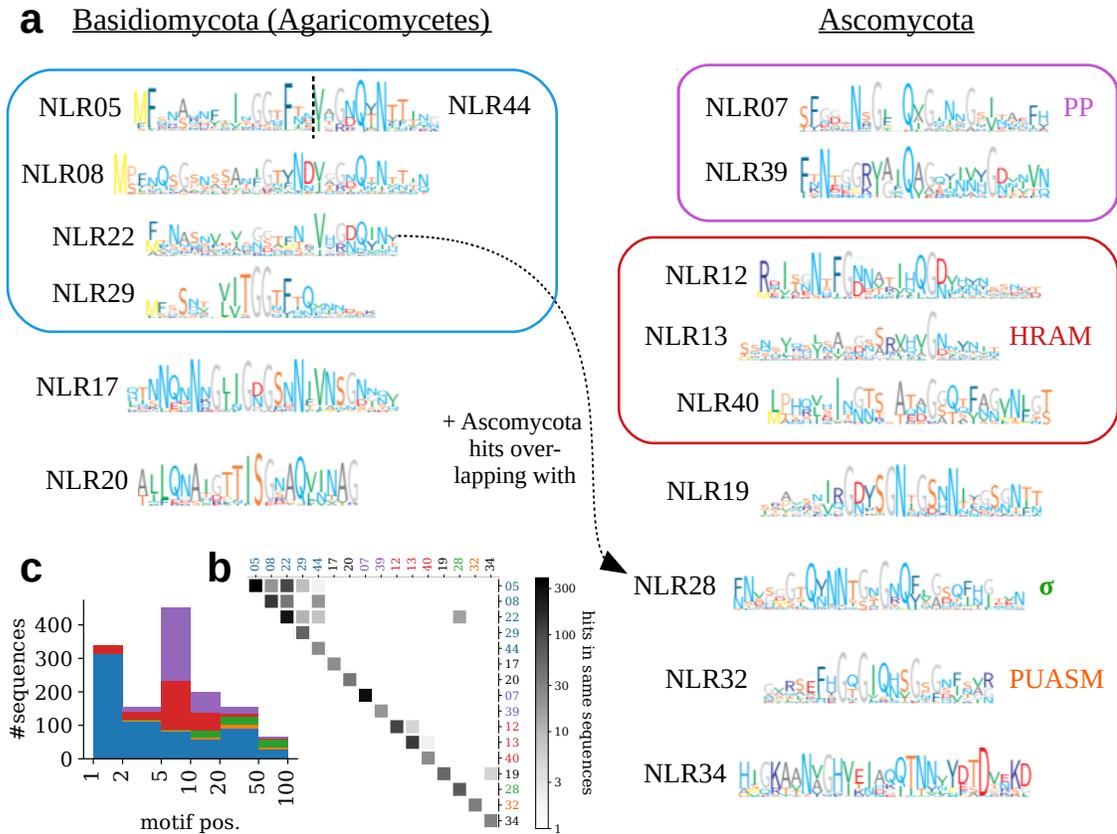


Figure 3: **Amyloid-like motifs in short N-termini of NLRs.** (a) Motif logos — grouped according to panel (b) Overlapping hits. See main text for details. (c) Stacked histogram of motif positions in NLR N-termini (the five main families of motifs color-coded as in panel (a)).

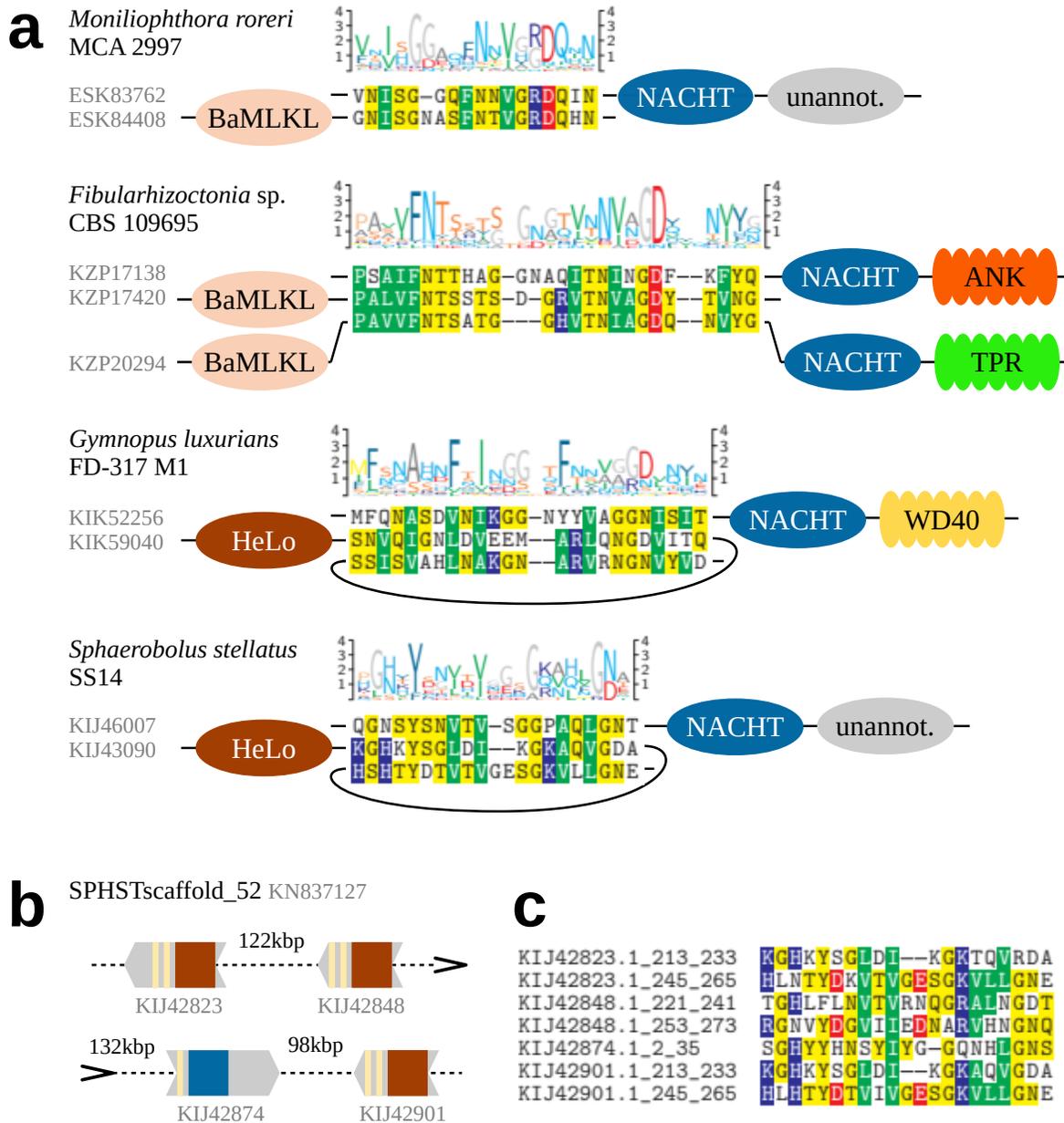
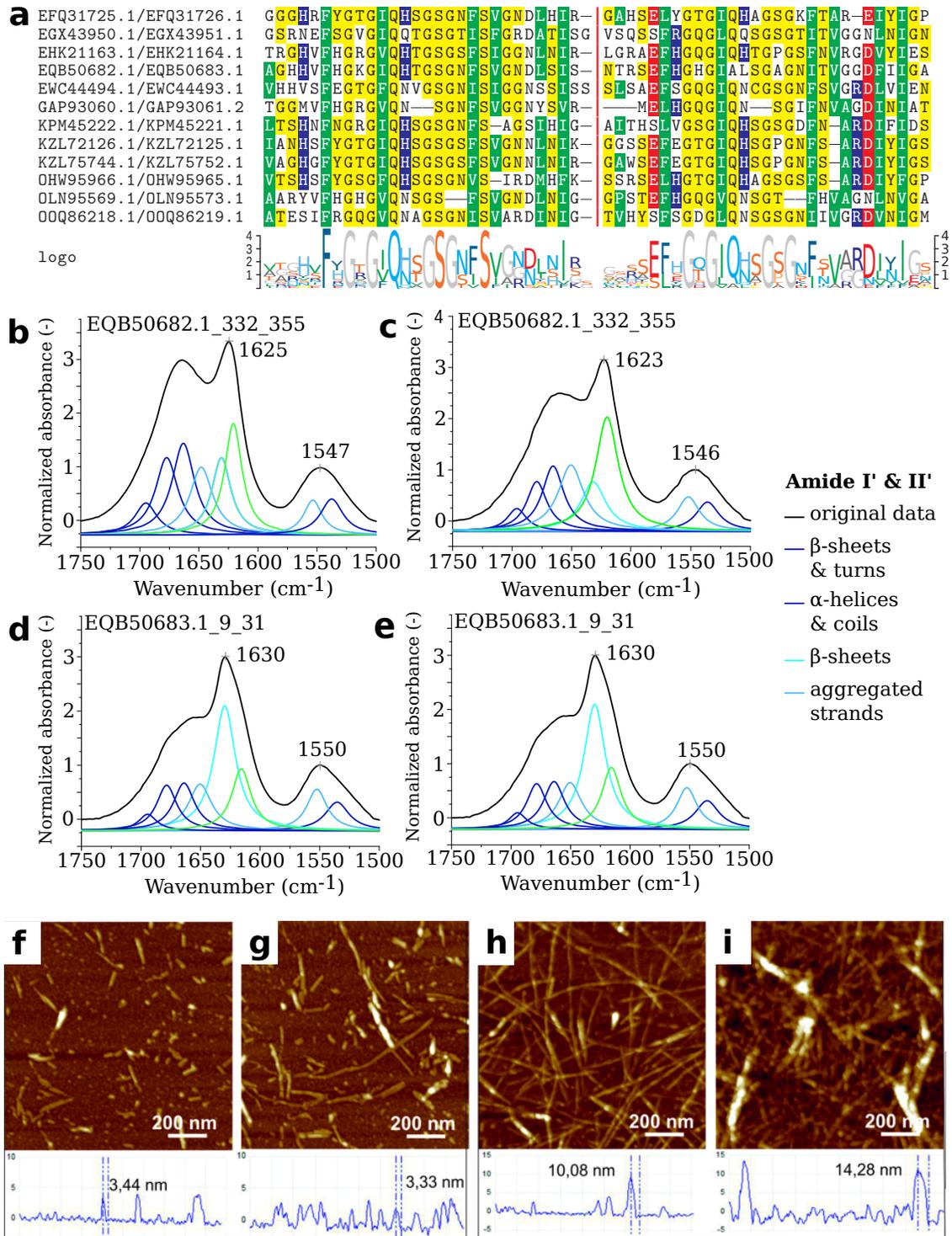


Figure 4: **Potentially interacting amyloid-like motifs in Agaricomycetes.** (a) Motif logos, sequence alignments and domain architectures of selected motif instances. (b) Schematic representation of a cluster of amyloid-like motifs in contig SPHSTscaffold\_52 from genome assembly GCA\_000827215.1 of *Sphaerobolus stellatus* SS14. (c) Multiple sequence alignment of motif instances in (b).



**Figure 5: The PUASM motif.** Panel **a**: Alignment of the PUASM motif pairs (left: C-terminal, right: N-terminal). Colors indicate residue hydrophobicity, curly brackets — the motif ranges. Panels **bcd**e: Normalized ATR-FTIR spectra of air-dried peptide films of EQB50682.1\_332\_355 (**bc**) and EQB50683.1\_9\_31 (**de**) with sub-bands obtained from the curve fitting procedure in the amide bands region (1750–1500  $\text{cm}^{-1}$ ) registered at 20°C. Panels **fghi**: AFM images with cross-section profiles of peptides EQB50682.1\_332\_355 (**fh**) and EQB50683.1\_9\_31 (**gi**). Samples were measured and imaged after dissolving (**bdfg**) and after 40 days of incubation at 37°C (98.6°F) (**cehi**).

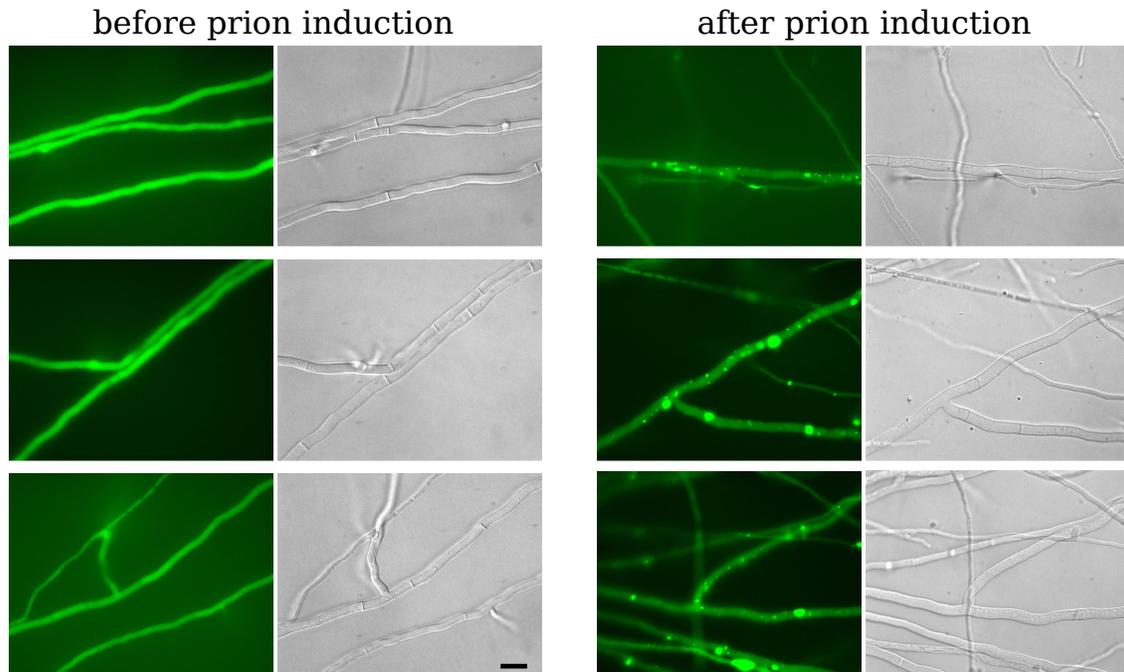


Figure 6: **Expression of GFP-fused PUASM motif in *Podospora anserina*.** Micrographs of *P. anserina* strains expressing molecular fusions of PUASM motif with GFP (in N- terminus) (Scale bar 5  $\mu$ m). Transformants displayed an initial diffuse fluorescence phenotype (**left** side of the panels) and acquired dot-like fluorescent aggregates, prion phenotype, after contact with strains already spontaneously expressing the infectious aggregated prion state (**right** side of the panels).