An extended β-finger motif is a near-universal feature of kinase fusion proteins

Oliver R. Powell[1], David Gilbert[1], Brande B. H. Wulff[1*]

[1]Plant Science Program, Biological and Environmental Science and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.

*Corresponding author. Email: brande.wulff@kaust.edu.sa

Abstract

Kinase fusion proteins (KFPs) are immune receptors conferring disease resistance in wheat and barley. We found that those with a confirmed function typically share an extended β-finger motif, a feature that originated ~98 million years ago in the Poaceae. Moreover, these functionally active KFPs are among the most highly expressed within a plant's full KFP repertoire, suggesting that elevated transcript abundance is a hallmark of disease-resistance activity.

Main Text

Kinase fusion proteins (KFPs) are a diverse set of immune receptors that confer resistance to a range of pathogens in wheat and barley. Previously, structural analysis of KFPs led to the identification of a conserved "extended" β-finger motif, uniquely present in the kinase domains of most KFPs [1,2]. In the Sr62-TK and Rwt4/Pm24 proteins the motif is involved in the recognition of the pathogen effectors AvrSr62 and Pwt4, respectively[1,3,4]. The motif also maintains the Sr62-TK homodimer required for function[1]. Therefore, the prevalence of this motif in KFPs with a validated disease resistance function suggests this feature may be suitable for the *ab initio* identification of KFPs.

To investigate the sequence conservation and diversity of the extended β-finger motif in KFPs, we extracted the sequences of the motif from AlphaFold3 predictions of 18 experimentally validated KFPs (Table S1). Yr15 was excluded, as the β-finger motif in both kinase domains lacked the extended conformation. These sequences were used to construct an initial hidden Markov model (HMM) of the motif. We iteratively refined the HMM using the proteome of bread wheat (*Triticum aestivum*) cv. Chinese Spring to improve its specificity and sensitivity (Fig. S1)[5].

This established a motif with a highly conserved length, although the amino acid sequences within the motif varied considerably across KFPs. Based on these structural predictions, we define the extended β-finger as a 31-amino-acid motif that extends the highly conserved 4th and 5th β-sheets of the kinase domain (Fig. 1A-B)[6].

Using the refined HMM for the extended β-finger motif, we annotated the proteome of Chinese Spring and identified 476 unique KFP loci across the genome. These loci were distributed across all chromosomes, with an enrichment in the gene-rich telomeric regions—similar to the distribution of nucleotide-binding leucine-rich repeat (NLR) resistance gene loci (Fig. 1C, Table S2)[7]. The number of KFPs per chromosome ranged from 6 on chromosome 4D to 53 on chromosome 7B (Table S3).

To investigate the evolutionary relationships among KFPs, we extracted the extended β-finger kinase domains from the protein sequences corresponding to the 476 unique KFP loci identified in Chinese Spring. These sequences were used to construct a phylogenetic tree, to which previously cloned KFPs were added for reference (Fig. 1D, Table S1).

We annotated the protein sequence of each KFP and identified 31 distinct auxiliary modules (i.e., fused domains) (Fig. 1D, Table S4). Only 10 of these 31 modules have previously been reported in cloned KFPs (Table S5) [2]. The most prevalent auxiliary modules included non β-finger kinases (*n* = 59), major sperm protein (MSP) domains (*n* = 53), MSP domains combined with WD40 repeats (MSP-WD40; *n* = 50) NBARC-LRR domains (*n* = 29) and coiled-coil NLRs (NLR; *n* = 25) (Table S5). Clades did not correlate with specific subgenomes (Fig. S2).

When mapped onto the KFP phylogeny, the auxiliary domains were found to cluster into distinct clades. In several cases, the same auxiliary modules appeared in multiple, independently evolved clades, suggesting convergent domain acquisition (Fig. S3). We also identified 150 kinase domains lacking detectable auxiliary domains (singletons), of which 55 were putatively annotated as cysteine-rich receptor-like kinases (crRLKs). Singletons formed several distinct clades, but sequences were also dispersed throughout the phylogenetic tree (Fig. S4).

We also identified 29 KFPs containing additional extended β-finger kinase domains, which clustered into clades associated with specific auxiliary modules (Fig. 1D) (Table S6). The most common auxiliary modules in these KFPs were kinase (*n* = 6), HMA-like (*n* = 6), and MSP-WD40 (*n* = 4) domains. In addition, we found 11 tandem kinase domains with additional β-finger motifs but no associated auxiliary domain.

The functionally validated KFPs $Sr62^{TK}$ and Rwt4/Pm24 trigger immune response via a shared CC-NLR, which is encoded by a locus located 20 kb distal and 150 kb proximal to the respective KFP-encoding genes. Therefore, we investigated whether other KFPs in Chinese Spring were also located in proximity to CC-NLRs across genomes and found that 28.2% of KFPs were located within 200 kb of a CC-NLR (Fig. S5, Table S7).

Most NLRs validated as resistance genes fall within the top 25% of the most highly expressed NLRs[8]. To assess whether KFPs conferring disease resistance also show elevated expression, we quantified the transcript levels of 12 KFPs using publicly available RNA-seq datasets (Table S8). We found that these previously cloned KFPs consistently ranked among the top 25% of expressed KFPs across each dataset except for *Rpg1* (Figure 1E, Figure S6A-I). Moreover, $Sr62^{TK}$ and $Sr62^{NLR}$ pair were expressed to similar levels and both in the top 25% of the most expressed KFPs and NLRs, respectively (Fig. S6H and S8.)

To investigate the evolutionary origin of the extended β-finger in plants, we annotated the motif in available reference genomes of monocot species. The extended β-finger motif was found exclusively in members of the *Poaceae* family, except for the early-diverging grass *Streptochaeta angustifolia* (Fig 2). This pattern suggests that the motif emerged at least 98 million years ago, following the divergence of *S. angustifolia* and *Pharus latifolius*[9].

Through computational analysis of cloned KFPs, we demonstrate that the extended β-finger is a highly conserved motif that can be used for the *de novo* annotation of KFPs. We believe this will be useful for the molecular cloning of novel resistance genes from Poaceae genomes, as KFPs comprise ~20% of currently cloned resistance genes in wheat[2]. Our analysis of the wheat

cultivar Chinese Spring genome identified 476 unique KFP loci; however, as most functionally validated KFPs are in the top 25% of expressed KFPs it is likely that not all the identified KFPs confer disease resistance[7,8].

Across the KFP phylogeny in Chinese Spring, distinct clades share the same auxiliary module. KFPs with MSP and kinase auxiliary modules form two and five distinct clades, respectively. The low sequence similarity between kinase domains in these clades suggests that KFPs in the different clades likely arose from distinct independent fusions, whereas those within a clade may share the same ancestral fusion event. Given their abundance and presence across many clades in the tree, crRLKs are likely the source of the kinase domains involved in these fusion events. This suggests that KFPs are the result of multiple independent fusion events. The identified singleton kinase domains may correspond to the pathogenicity targets of the effectors recognized by the KFPs in these clades[10]. Alternatively, the auxiliary modules fused to the kinases may represent integrated decoys for homologous effector targets similar to NLRs with integrated domains [11].

Although KFPs have been cloned exclusively from barley, wheat, and wheat wild relatives, our phylogenetic analysis of monocot proteomes highlights a potentially untapped wealth of KFP resistance available in members of the Poaceae such as rice, maize, and their wild relatives. Our analysis also highlights that the extended ß-finger motif is private to Poaceae. However, since crRLKs are ubiquitous across plant species, it is possible that KFPs lacking the extended ß-finger motif exist outside the Poaceae. This may also explain the absence of the motif in the kinase domains of Yr15 [2].

Materials and Methods

Extended β-finger HMM profile development

AlphaFold3 was used to predict the structures of cloned KFPs (Table 1). Structures were visualised in PyMol and manually curated to identify and extract the sequence of the extended β-finger motif. A multiple sequence alignment (MSA) of unique extended β-finger sequences was generated using Muscle v5.1 [12]. The MSA was used to create a seed HMM using *hmmbuild* with default parameters [13].

The seed HMM was then iteratively refined on the Chinese Spring reference proteome. In each iteration step, *hmmsearch,* with parameter *–incdome 1e-10* and *-A proteome.aln*, was used to identify and align β-finger containing protein sequences [13]. The number of unique sequences was counted and, using the alignment, a new HMM was constructed using *hmmbuild*. The newly created HMM was then used for the next iteration step [13]. This continued until three continuous iteration steps identified the same number of proteins in the proteome. The number of sequences annotated in each refinement step was plotted using plot_refinement.py (https://github.com/Orpowell/wheat_kfps). The matrix for the refined HMM was extracted using Skylign and the sequence logo of the HMM was plotted plot_sequence_logo.py (https://github.com/Orpowell/wheat_kfps) [14].

Chinese Spring annotation and phylogenetics

A proteome of the telomere-to-telomere (T2T) Chinese Spring assembly was generated by selecting the longest protein sequence at each annotated locus using get_locus_representatives.py (https://github.com/Orpowell/wheat_kfps) [15]. The reduced proteome was annotated using *hmmsearch,* with the parameter --domE 1e-4, and the extended β-finger HMM (see above) and the Pkinase HMM from PFAM [13,16]. A protein sequence was classified as a KFP if it contained at least one overlapping extended β-finger and Pkinase annotation.

The primary module, extended β-finger kinase domain(s), of Chinese Spring KFPs and previously cloned KFPs were extracted using extract_kfp_sequences.py (https://github.com/Orpowell/wheat_kfps), and aligned using MUSCLE v5.1 with default parameters [12]. The resulting multiple sequence alignment was used to create a phylogenetic tree using FastTree (v2.1.11) with default parameters [17].

The phylogenetic tree was visualized using iTOL v7 [18]. An iTOL annotation for the number of β-fingers in each KFP was generated using count_beta_fingers.py (https://github.com/Orpowell/wheat_kfps). Protein sequences of Chinese Spring KFPs and previously cloned KFPs (Table S1) were annotated using InterProScan, with TMHMM enabled [19,20]. The auxiliary module of each KFP was identified and used to create an iTOL annotation using annotate_KFP_auxiliary_domain.py (https://github.com/Orpowell/wheat_kfps). The coordinates of KFP-encoding loci were extracted from the Chinese Spring reference GFF file and visualised using plot_KFP_physical_positions.py (https://github.com/Orpowell/wheat_kfps). The subgenome of each KFP was extracted and annotated using annotate_subgenome.py (https://github.com/Orpowell/wheat_kfps).

Chinese Spring KFP-NLR proximity

NLR loci in the T2T Chinese Spring genome were annotated using NLR-Annotator with default parameters [21]. The nearest CC-NBARC-LRR locus to each KFP locus (identified above) was determined using calculate_nlr_proximity.py (https://github.com/Orpowell/wheat_kfps).

HMM-Annotator

HMM-Annotator is a Python tool for the annotation of genomic and transcriptomic sequences using protein HMMs. The tool works in the following steps. First the input sequence is dissected into overlapping fragments. Each fragment is then fully translated into 6 protein sequences (three forward and three reverse open reading frames) that are annotated for HMM using PyHMMer. Protein sequence annotations are then converted back to their original nucleic acid sequence coordinates. The output can be saved as a TSV or a BED file. HMM-Annotator was written in Python (v3.10.14). Source code and further documentation can be found on GitHub (https://github.com/Orpowell/HMM-Annotator).

Cloned KFP transcript expression analysis

Publicly available RNA-seq data for cloned KFPs was downloaded from NCBI and NGDC (Table S). The raw reads from each dataset were trimmed to remove adapters and low-quality sequences using trimmomatic (v.0.39) with parameters: LEADING:5,TRAILING:5, SLIDINGWINDOW:4:20, MINLEN:50 [22]. Trimmed reads were *de novo* assembled using Trinity (v2.15.2) using default parameters [23]. Transcript abundance was estimated using Salmon (v1.10.3), with bias correction enabled (--gcBias, --seqBias), and aggregated into gene-level TPMs using the R package tximport (v1.36.0) [24,25]. β-finger containing transcripts were identified using HMM-Annotator (as described above), and gene TPMs were filtered for β-finger containing transcripts. Transcripts corresponding to cloned KFPs were identified using BLASTn [26]. KFP gene expression data was visualised using TEA-tools visualise command (https://github.com/Orpowell/tea-tools). Analysis for Sr62-NLR was performed as described above. The GenBank sequence, PP537390.1, was used for annotation of Sr62-NLR transcript.

KFP annotation of monocots

Publicly available reference genomes of monocot species were downloaded from NCBI using the NCBI datasets command: "datasets download genome taxon 4447 –dehydrated –reference –include genome". The genome of Streptochaeta angustifolia was downloaded using the NCBI datasets command: "download genome accession GCA_020804685.1 –include genome". The name of each species was extracted from the "assembly_report.jsonl" included in the data downloads using get_species_name.py (https://github.com/Orpowell/wheat_kfps). A phylogenetic tree of species was downloaded from NCBI using ete3 and visualised using iTOL v7 [18,27].

Each genome was annotated for extended β-finger motifs using HMM-Annotator and the extended β-finger motif HMM (see above). The output BED file for each genome was filtered for annotations with a minimum length of 90 nucleotides and score (E-value) less than 1E-10. Presence or absence of the motif was determined from the filtered BED file and used to generate an iTOL annotation for the phylogenetic tree. All analysis was performed using generate_itol_annotations.py (https://github.com/Orpowell/wheat_kfps).

Acknowledgements:
This research used the Ibex HPC cluster managed by the Supercomputing Laboratory at King Abdullah University of Science and Technology (KAUST). We are grateful to Burkhard Steuernagel for helpful discussions during the development of HMM-Annotator.

Funding:
This research was funded by KAUST baseline and award CRG11-2022-5087 to B.B.H.W.

Author Contributions:
O.R.P. developed the β-finger HMM and HMM-Annotator. O.R.P. performed all bioinformatic analysis related to Chinese Spring and the evolutionary origin of the extended β-finger motif. O.R.P. developed HMM-Annotator and TEA-tools. D.G. performed expression analysis of cloned KFPs. B.B.H.W. and O.R.P. conceived the study. O.R.P. drafted the manuscript with contributions from B.B.H.W and D.G. All coauthors read and approved the final manuscript.

Data and Materials Availability:
Source code for HMM-Annotator, Tea-tools and all analysis are available on GitHub: (https://github.com/Orpowell/HMM-Annotator), (https://github.com/Orpowell/tea-tools), (https://github.com/Orpowell/wheat_kfps). The HMM for the extended β-finger motif is archived on Zenodo (https://doi.org/10.5281/zenodo.15874664).

Fig 1. Annotation and expression of extended β-finger KFPs in bread wheat.
a, The extended β-finger kinase domain of Sr62-TK (Table S1). **b**, Sequence logo of the refined HMM profile representing the extended β-finger in KFPs from wheat (*Triticum aestivum*) cv. Chinese Spring. **c**, Genomic distribution of KFP-associated loci across the Chinese Spring genome. **d**, Phylogenetic tree of KFP protein sequences in Chinese Spring, constructed from the primary module (i.e., extended β-finger domain[s]) of each KFP. Black circles indicate bootstrap value ≥80. Previously cloned KFPs were included and labelled. Sequences are coloured according to their number of extended β-finger kinase domains (inner circle) and auxiliary module (outer circle), as indicated in the figure legend. Auxiliary domains occurring less than 3 times across the phylogenetic tree are labelled miscellaneous (Table S4-5). **e,** KFP transcript abundance in *de novo* assembled transcriptomes of *T. aestivum* cv. Cadenza. Functionally validated KFPs cloned from each species are highlighted. Gene transcript levels

were estimated from self-aligned RNA-seq data expressed as transcripts per million (TPM). The top 25% most expressed KFP transcripts are coloured orange.

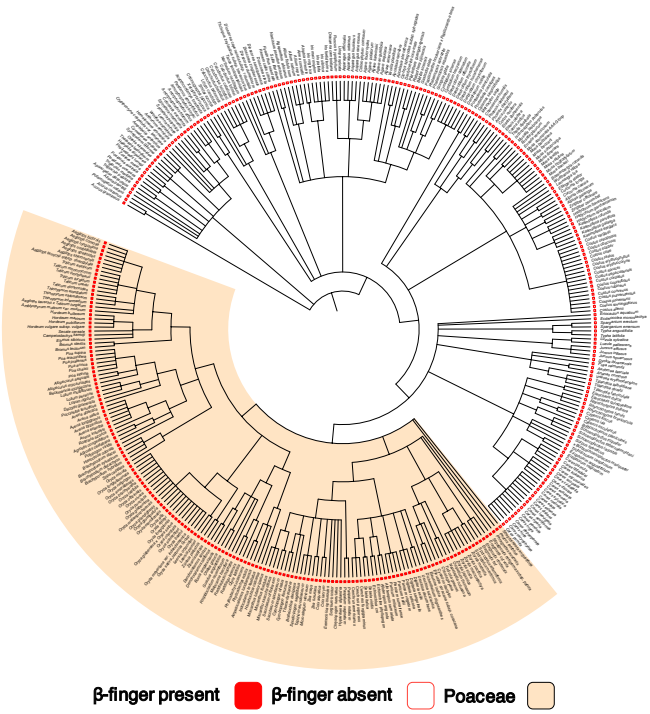


Fig 2. Extended β-finger KFPs are private to the Poaceae family.

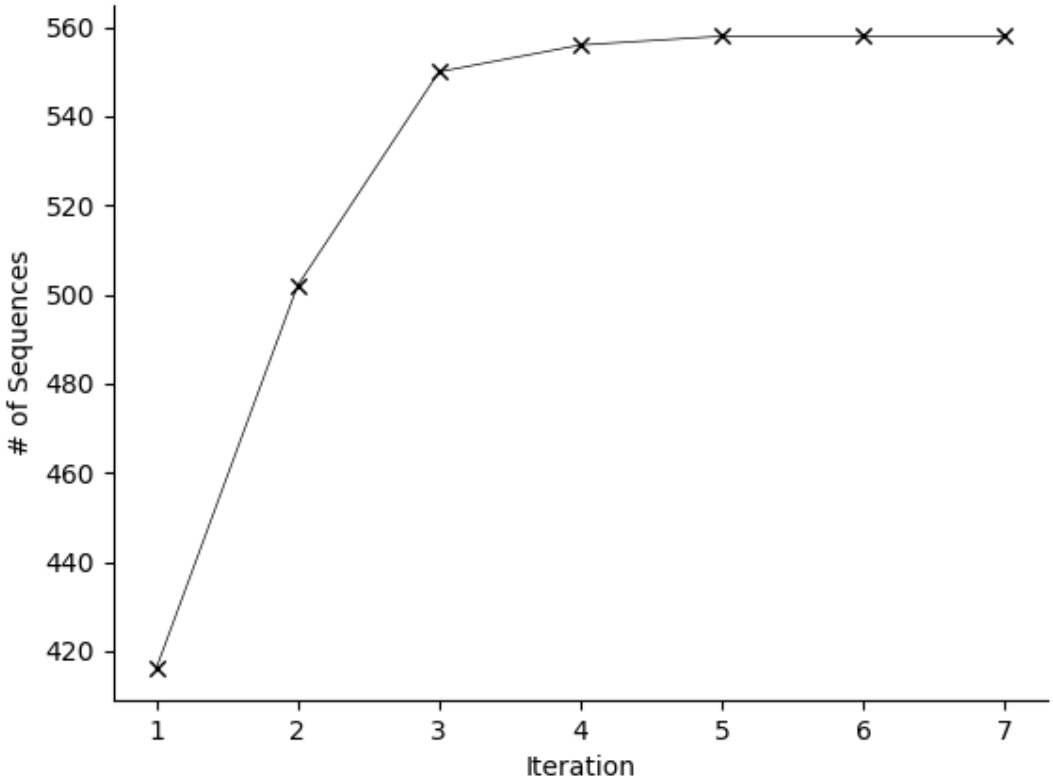
The presence of the extended β-finger motif in publicly available genomes of monocotyledonous species. A solid red square indicates species in which the motif was detected. A white square with a red outline indicates species in which the motif was not detected. Members of the family Poaceae are coloured beige.

References

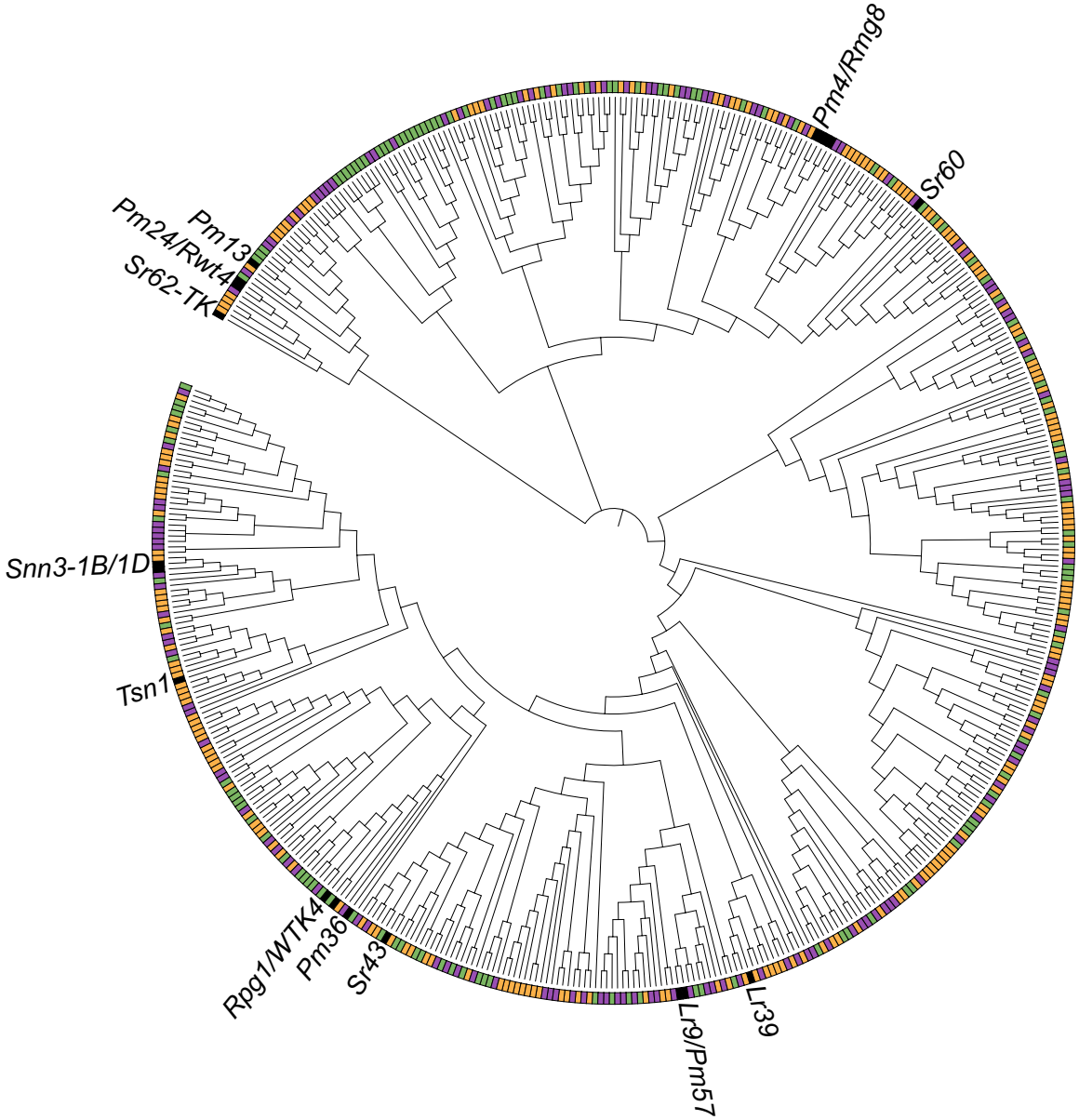
1. Chen, R. *et al.* A wheat tandem kinase activates an NLR to trigger immunity. *Science (1979)* **387**, 1402–1408 (2025).
2. Powell, O. R. *et al.* Kinase fusion proteins bolster cereal immunity. *In Submission* (2025).
3. Sung, Y. C. *et al.* Wheat tandem kinase RWT4 directly binds a fungal effector to activate defense. *Nature Genetics 2025 57:5* **57**, 1238–1249 (2025).
4. Lu, P. *et al.* A wheat tandem kinase and NLR pair confers resistance to multiple fungal pathogens. *Science* **387**, 1418–1424 (2025).
5. Liu, S. *et al.* A telomere-to-telomere genome assembly coupled with multi-omic data provides insights into the evolution of hexaploid bread wheat. *Nature Genetics 2025 57:4* **57**, 1008–1020 (2025).
6. Hanks, S. K., Quinn, A. M. & Hunter, T. The Protein Kinase Family: Conserved Features and Deduced Phylogeny of the Catalytic Domains. *Science (1979)* **241**, 42–52 (1988).
7. Steuernagel, B. *et al.* The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor Repertoire. *Plant Physiol* **183**, 468–482 (2020).
8. Brabham, H. J. *et al.* Discovery of functional NLRs using expression level, high-throughput transformation, and large-scale phenotyping. *bioRxiv* 2024.06.25.599845 (2024) doi:10.1101/2024.06.25.599845.
9. Ma, P. F. *et al.* The Pharus latifolius genome bridges the gap of early grass evolution. *Plant Cell* **33**, 846–864 (2021).
10. Wan, C. *et al.* A serine-rich effector from the stripe rust pathogen targets a Raf-like kinase to suppress host immunity. *Plant Physiol* **190**, 762–778 (2022).
11. Marchal, C., Michalopoulou, V. A., Zou, Z., Cevik, V. & Sarris, P. F. Show me your ID: NLR immune receptors with integrated domains in plants. *Essays Biochem* **66**, 527 (2022).
12. Edgar, R. C. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nature Communications 2022 13:1* **13**, 1–9 (2022).
13. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195 (2011).
14. Wheeler, T. J., Clements, J. & Finn, R. D. Skylign: A tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics* **15**, 1–9 (2014).
15. Liu, S. *et al.* A telomere-to-telomere genome assembly coupled with multi-omic data provides insights into the evolution of hexaploid bread wheat. *Nature Genetics 2025 57:4* **57**, 1008–1020 (2025).
16. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412–D419 (2021).
17. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **5**, e9490 (2010).
18. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* **52**, W78–W82 (2024).
19. Paysan-Lafosse, T. *et al.* InterPro in 2022. *Nucleic Acids Res* **51**, D418–D427 (2023).

20. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol* **305**, 567–580 (2001).
21. Steuernagel, B. *et al.* The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor Repertoire. *Plant Physiol* **183**, 468–482 (2020).
22. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114 (2014).
23. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc* **8**, 10.1038/nprot.2013.084 (2013).
24. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Res* **4**, (2016).
25. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods 2017 14:4* **14**, 417–419 (2017).
26. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 1–9 (2009).
27. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol* **33**, 1635–1638 (2016).

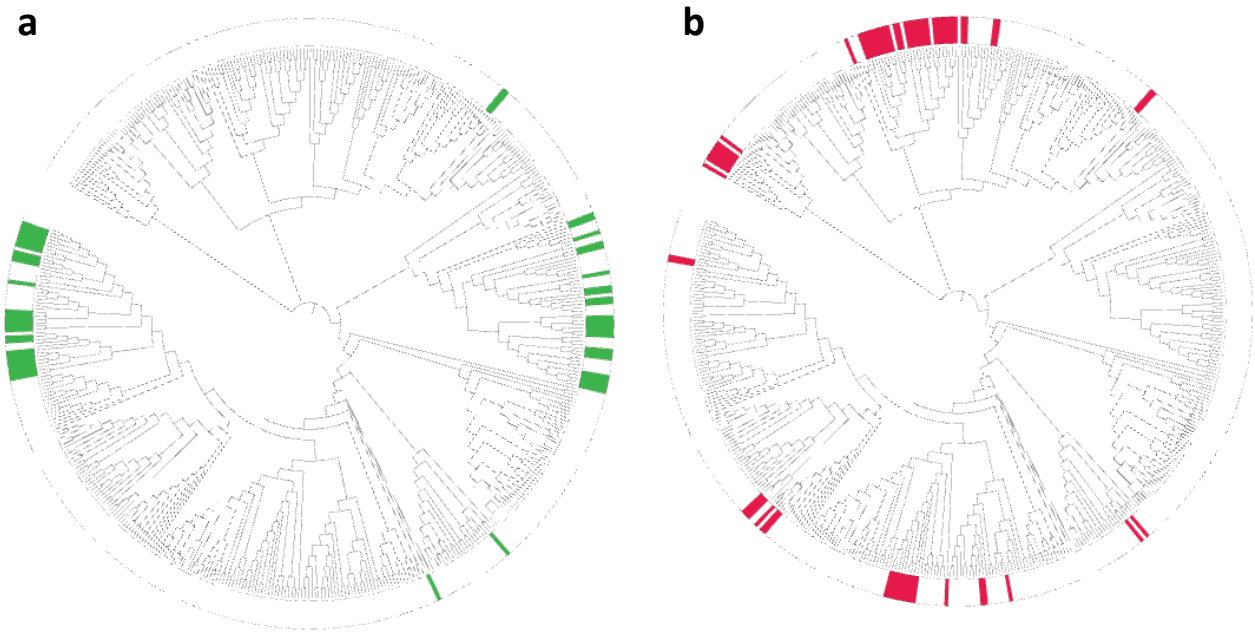
Supplementary Figures



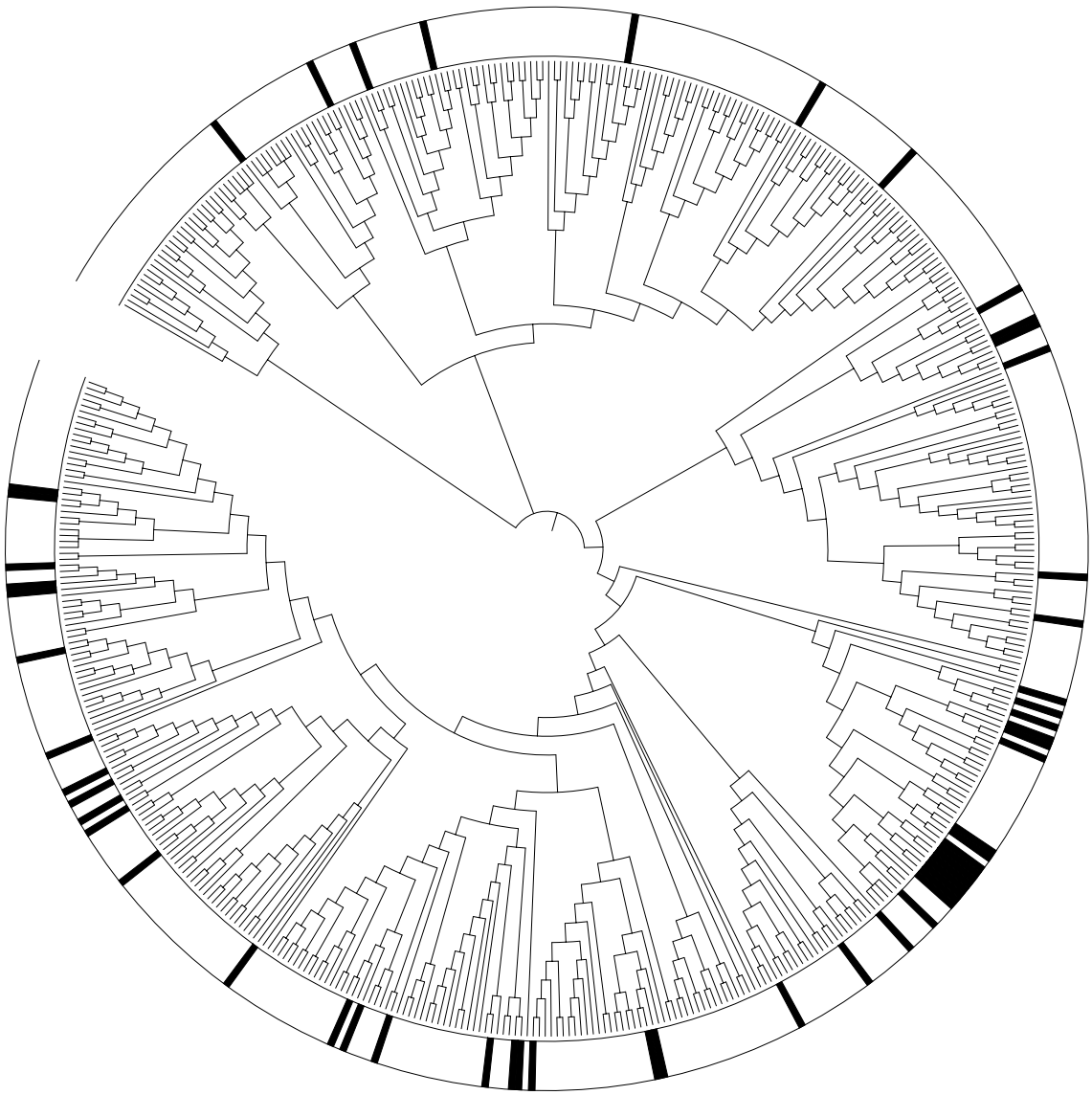
Supplementary Fig. 1. Iterative refinement of the extended β-finger HMM.
The seed HMM of the extended β-finger motif, derived from sequences of cloned KFPs, was iteratively refined on the Chinese Spring proteome until three consecutive iteration steps annotated the same number of sequences in the proteome.



Supplementary Fig. 2. Subgenome distribution of Chinese Spring KFPs.
Phylogenetic tree of Chinese Spring KFPs. The subgenome origin of each KFP is coloured according to the legend. Previously cloned KFPs are coloured black and labelled.

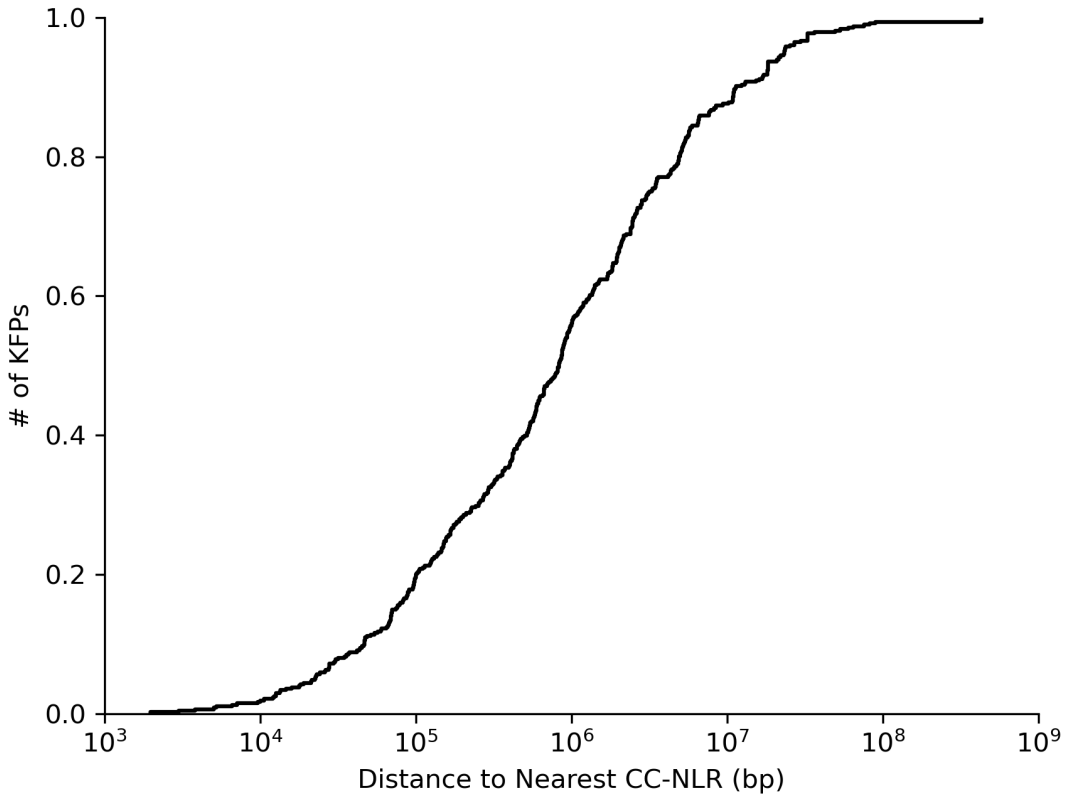


Supplementary Fig. 3. Distribution of auxiliary modules across a phylogenetic tree of Chinese Spring KFPs.
The distribution of MSP (**a**) and kinase (**b**) auxiliary modules across the phylogenetic tree of Chinese Spring KFPs.

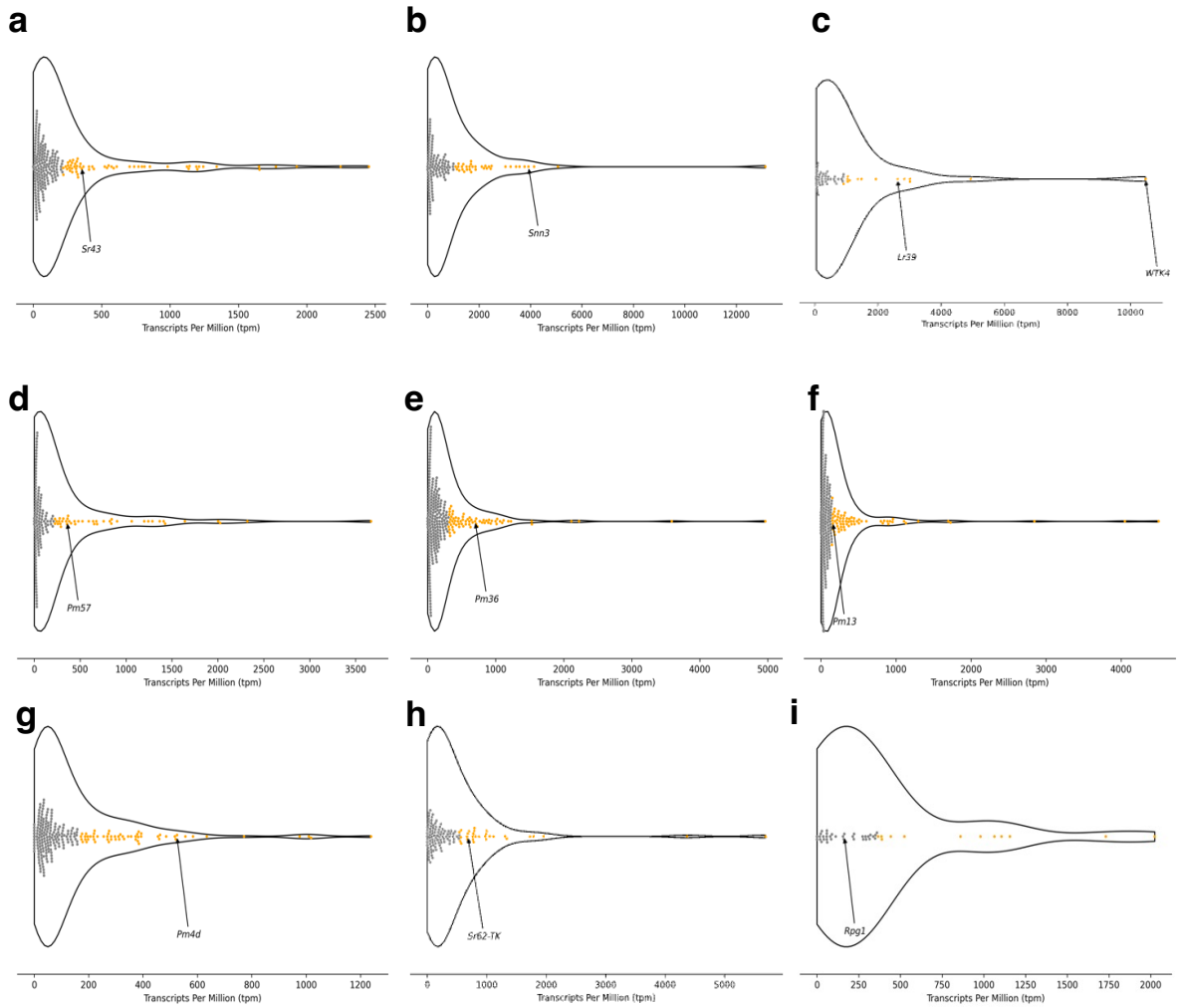


Supplementary Fig. 4. Distribution of extended β-finger kinases across a phylogenetic tree of Chinese Spring KFPs.

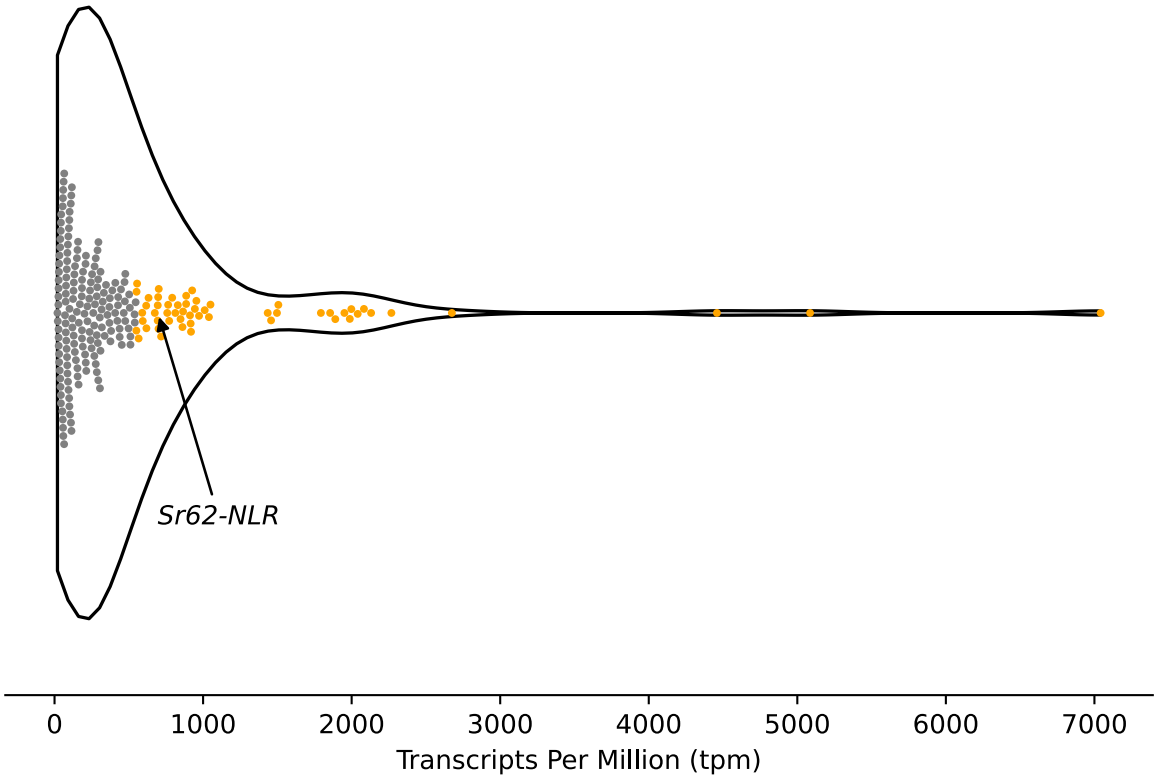
Phylogenetic tree of Chinese Spring KFPs. The positions of extended β-finger kinases, *i.e.* kinases with an extended β-finger but no fused auxiliary module, are coloured black.



Supplementary Fig. 5. Distance to nearest CC-NLR for each KFP in Chinese Spring. Empirical distribution function of the distance of Chinese Spring KFPs (n=476) to the nearest CC-NLR locus. Approximately 28.2% of KFPs were located within 200 kb of a CC-NLR.



Supplementary Fig. 6. Functional KFPs are the most expressed KFP transcripts. Gene abundance of KFPs in *de novo* assembled transcriptomes of *Thinopyrum elongatum* (**a**), *Aegilops tauschii* (**b, c**), *Ae. searsii* (**d**), *T. turgidum* ssp. *dicoccoides* (**e**), *Ae. longissima* (**f**), *T. turgidum* (**g**) *Ae. sharonensis* (**h**) and *Hordeum vulgare* (**i**). Functionally validated KFPs cloned from each species are highlighted. Gene transcript levels were estimated from self-aligned RNA-seq data expressed as transcripts per million (TPM). The top 25% most expressed transcripts are coloured orange.



Supplementary Fig. 7. Sr62-NLR is highly expressed among NLR transcripts.
Gene abundance of NLRs in the *de novo* assembled transcriptome of *Aegilops sharonensis*. The functionally validated helper NLR of Sr62-TK, Sr62-NLR is labelled. Gene transcript levels were estimated from self-aligned RNA-seq data expressed as transcripts per million (TPM). The top 25% most expressed transcripts are coloured orange.

Supplementary Tables S1 to S8 are provided in a separate excel file as Data S1

Table S1. Genbank IDs of cloned kinase fusion proteins.

Table S2. Physical position of kinase fusion froteins in Chinese Spring.

Table S3. Number of KFPs per chromosome in Chinese Spring.

Table S4. Auxiliary modules of KFPs in Chinese Spring.

Table S5. Number of KFPs with each auxiliary module.

Table S6. Number of β-fingers in each KFP.

Table S7. Distance to Nearest CC-NLR for each KFP.

Table S8. RNA-Seq datasets for functionally validated KFPs.