nature genetics

Analysis

https://doi.org/10.1038/s41588-024-02032-x

Tandem kinase proteins across the plant kingdom

Received: 23 May 2024

Accepted: 11 November 2024

Published online: 8 January 2025

Check for updates

Tamara Reveguk[1,2], Andrii Fatiukha[1,2,3], Evgenii Potapenko[1,2], Ivan Reveguk[4], Hanan Sela[1], Valentyna Klymiuk[1,2,3], Yinghui Li[1,2,5], Curtis Pozniak[3], Thomas Wicker[6], Gitta Coaker[7] ✉ & Tzion Fahima[1,2] ✉

Plant pathogens pose a continuous threat to global food production. Recent discoveries in plant immunity research unveiled a unique protein family characterized by an unusual resistance protein structure that combines two kinase domains. This study demonstrates the widespread occurrence of tandem kinase proteins (TKPs) across the plant kingdom. An examination of 104 plant species' genomes uncovered 2,682 TKPs. The majority (95.6%) of these kinase domains are part of the receptor-like kinase–Pelle family, which is crucial for cell surface responses in plant immunity. Notably, 90% of TKPs comprise dual kinase domains, with over 50% being pseudokinases. Over 56% of these proteins harbor 127 different integrated domains, and over 47% include a transmembrane domain. TKP pseudokinases and/or integrated domains probably serve as decoys, engaging with pathogen effectors to trigger plant immunity. The TKP Atlas we created sheds light on the mechanisms of TKP convergent molecular evolution and potential function.

Plants are essential for food production. Unfortunately, crop production is impacted by various biotic stresses[1]. The perpetual interaction of plants and pathogens is evident in their defense and counterdefense strategies[2]. Plant immune receptors capable of recognizing and responding to pathogen attack include surface-localized pattern recognition receptors (PRRs) and intracellular nucleotide-binding leucine-rich repeat (NLR) receptors[3].

Surface-localized PRRs can recognize a variety of biological patterns present in pathogens as well as extracellular pathogen effector proteins[3,4]. The patterns are called pathogen-associated molecular patterns or microbial-associated molecular patterns and include flagellin, lipopolysaccharides and chitin[3]. PRRs are partitioned into receptor-like kinases (RLKs) and receptor-like proteins (RLPs)[3]. RLKs comprise the following three domains: an extracellular ligand-binding domain, a transmembrane domain and an intracellular protein kinase domain (KD)[5]. RLPs, in contrast, while having extracellular and transmembrane domains, lack a KD within their short cytoplasmic region[6].

Plant pathogens modulate plant metabolism and immune responses to their benefit by secreting effector proteins into plant cells. To combat this, plants employ numerous intracellular NLR receptors capable of recognizing effectors and mounting effector-triggered immunity[7]. NLR activation induces programmed cell death, also called the hypersensitive response[7]. NLRs can directly detect pathogen effectors or indirectly detect modifications of host proteins[8]. Notably, some NLRs, known as NLR-integrated domain (NLR-ID) receptors, incorporate both strategies within their structure—direct via integrated domain and indirect recognition by host protein monitoring[8].

Additionally, NLRs can interact with host kinases. The ZAR1 NLR, for example, interacts with the effector target, pseudokinase ZED1, and is capable of identifying several pathogen effectors[9]. Downstream signaling of PRRs and NLRs overlaps and mutually potentiates each other, despite having differing structural characteristics and identifying different pathogen components[10]. The activation of plant immune receptors triggers a series of responses: release of calcium, an elevation

[1]Institute of Evolution, University of Haifa, Haifa, Israel. [2]Department of Evolutionary and Environmental Biology, University of Haifa, Haifa, Israel. [3]Crop Development Centre and Department of Plant Sciences, University of Saskatchewan, Saskatoon, Saskatchewan, Canada. [4]Laboratory of the Structural Biology of the Cell (BIOC), École Polytechnique, Paris, France. [5]Triticeae Research Institute, Sichuan Agricultural University, Chengdu, China. [6]Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland. [7]Department of Plant Pathology, University of California, Davis, CA, USA. ✉e-mail: glcoaker@ucdavis.edu; tfahima@evo.haifa.ac.il

in reactive oxygen species, the activation of MAP kinases and alterations in gene transcription geared towards defense[11].

The discovery of the barley stem rust resistance gene *Rpg1* and the wheat stripe rust resistance gene *Yr15* (*WTK1*) prompted a new chapter in plant immunology[12,13]. Both encode tandem kinase proteins (TKPs) with KD–pseudokinase domain (KD–PKD) architecture. So far, more than ten cereal TKPs were functionally validated[12–22]. Two of the wheat tandem kinases (WTKs) contain integrated domains annotated as von Willebrand factor A (vWA)—*Lr9* (*WTK6-vWA*) governing strong resistance to leaf rust and *Pm57* (*WTK6b-vWA*) conferring resistance to powdery mildew[20,22]. The mutation analysis suggests that pathogen perception by TKPs can also involve integrated domains[20]. The wheat blast resistance gene *RWT4* specifically recognizes the *Magnaporthe oryzae* effector AvrPWT4, and its kinase activity is essential for defense[23]. Taken together, TKPs emerged as new players in plant immunity.

An initial search of 11 plant genomes, mostly cereals, aiming to explore TKP evolution, discovered 92 putative TKPs[12]. The study in ref. 12 proposed that multiple KDs originated via duplication or fusion, with over half of the predicted kinases likely resulting from gene duplication. However, no systematic attempts were made to explore TKP distribution across a broad range of plant species throughout the plant kingdom or to provide a comprehensive characterization of their structure, origin and potential functional role in plant immunity.

In the present study, we investigated the prevalence and domain architecture of TKPs in 104 genomes across the plant kingdom. Over half of the discovered TKPs contained a PKD, while 55.9% of all TKPs contained a nonkinase integrated domain. This suggests that TKP may use their pseudokinases or integrated domains as decoys to trap pathogen effectors. The discovery of the TKP protein family as an essential player in the plant immune system may reconceptualize our understanding of immune receptor activation in plants. Our TKP Atlas will serve as a base for future discoveries of TKP convergent molecular evolution and function.

Results

Identification of TKPs across the plant kingdom

To date, ten experimentally verified tandem kinases are known (Supplementary Table 1). To assess the prevalence of TKPs in the plant kingdom, we searched 104 plant genomes for the presence of protein kinases. In our analysis, we incorporated species from the following three classes: Monocotyledoneae, Eudicotyledoneae and Magnoliidae, spanning ploidy levels from diploid to hexaploid. We particularly focused on the 47 agricultural species and model plants like *Arabidopsis* and *Nicotiana benthamiana*. TKPs represent a protein family due to their similar structural composition, identified by the presence of two or more fused KDs. Therefore, we identified proteins that contain at least one KD by using the ProSITE kinase profile (PS50011). A total of 1,78,376 protein kinases were identified, constituting 3.8% of all analyzed proteins.

Next, we identified TKPs from the protein kinases based on the presence of two or more fused KDs. In our study of 104 plant genomes, we found that the number of TKPs per genome ranged from 1 (*Saccharum officinarum x spontaneum*) to 346 (*Vaccinium corymbosum*; Fig. 1). In total, we identified 2,682 proteins as TKPs, with the number of KDs per gene ranging from two to eight. However, the maximum number of KDs in confirmed multikinase proteins reached only five throughout our literature analyses[24]. Therefore, we removed six sequences from the atlas that contained more than five kinase/pseudokinases domains (PKDs), as they are likely annotation errors, leaving 2,676 TKPs in the subsequent analysis. Plant species differed in the number and composition of tandem kinases found in their genomes. The northern highbush blueberry (*V. corymbosum*) contained the largest number of TKPs in our analyses, with 346, which is twice the amount found in William's lovegrass (*Eragrostis tef*, 143 TKPs). However, *Leersia perrieri* had the highest percentage of TKPs relative to full proteome (0.35%), followed by *E. tef* (0.34%; Extended Data Fig. 1a). The number of TKPs weakly correlated with genome size (Spearman's correlation coefficient (ρ) = 0.232, P = 0.018), and no significant differences between sequencing technologies were found (ANOVA, P = 0.954; Supplementary Table 2).

We also analyzed the TKPs for the presence of integrated non-KDs. We identified that 1,180 protein sequences contained only KDs in tandem, while 1,496 proteins included integrated domains. These integrated domains fused to kinases are mostly associated with plant defense systems, such as lectin, leucine-rich repeat and stress–antifungal domains[25–28].

We used HMMER with a collection of KD profiles to refine KD membership. This analysis classified the 5,723 KDs into ten groups comprising 77 subfamilies (Supplementary Table 3). The RLK–Pelle group was the most prominent, with 5,472 members (95.6% of the total) partitioned into 46 subfamilies. Moreover, the largest subfamilies were RLK–Pelle_DLSV (25.3%) and RLK–Pelle_L-Lec (11.7%; Extended Data Fig. 1b). The DLSV family comprises the following four moss-specific RLKs subfamilies: DUF26, LRR-VIIIb, SD-1 and vWA[29,30]. The experimentally validated TKPs primarily belonged to the DLSV family in the LRR-VIIIb subfamily (Supplementary Table 1).

Prevalence of TKPs among cereal crops

Our analysis included eight species from the Triticeae tribe exhibiting a range of ploidy levels from 2× to 6×. Likely due to genome duplication events, polyploid species (4×, 6×) harbored several times more TKPs than diploid species (Supplementary Table 4).

Despite the difference in TKP counts, they are distributed on all chromosomes and spread across all three subgenomes (A, B and D), TKPs frequently clustered at the ends of chromosomes, as seen in the cases of *Aegilops tauschii, Triticum urartu*, *Triticum turgidum*, *Triticum aestivum* and *Triticum spelta* (Extended Data Fig. 2). Other resistance genes (that is NLRs) also show clusters near telomeres, where higher recombination rates drive faster evolutionary adaptations[31,32].

Variety of kinase families among TKPs

To assess the functionality of the identified tandem kinases, we categorized their domains into 'likely protein kinase' and 'pseudokinase' based on the presence of the following three catalytically important residues from the catalytic triad: β_3-lysine, HRD_{Asp} and DFG_{Asp}[33](Fig. 2a,b).

Among the analyzed TKPs, 2,372 (90%) are dual kinases. These TKPs exhibited various combinations of kinase and pseudokinase organization. The most prevalent combination was the presence of two KDs, which accounted for 42% (1,002) of all dual kinases. Combinations of KD–PKD and PKD–KD were almost as common, comprising 41% (989) of the analyzed dual kinases. The remaining 18% (381) of dual kinases TKPs contained two PKDs, as illustrated in Fig. 2c. Among the identified TKPs, combinations with two KDs from RLK–Pelle_DLSV (20%) or RLK–Pelle_L-Lec (10%) were the most frequent, with most domains categorized as functional (Fig. 2d). In summary, most TKPs from across 104 plant genomes contain a combination of two protein KDs (KD–KD or KD–PKD), which belong to the RLK–Pelle_DLSV family or RLK–Pelle_L-Lec family. Some compositions (e.g., CrRLK1L-1—RLCK-VIIa-2, CrRLK1L-1—DLSV) were present only in the species from Eudicotyledoneae clade (Fig. 2d).

Integrated non-KDs are common in TKPs

We analyzed the TKPs for the presence of integrated domains by scanning against available Pfam HMM profiles. Among the 2,676 TKPs, 1,496 proteins included integrated domains. These integrated domains, such as lectins, leucine-rich repeats and stress–antifungal domains, fused to kinases, are mostly associated with plant defense systems[25,28]. We also identified 1,267 (47.3%) TKPs that possess a transmembrane domain, a typical structural component of plant RLKs. Notably, 1,132 (89.3%) of these 1,267 TKPs contained both an integrated domain and a transmembrane domain. Numerous species displayed a considerable

https://doi.org/10.1038/s41588-024-02032-x

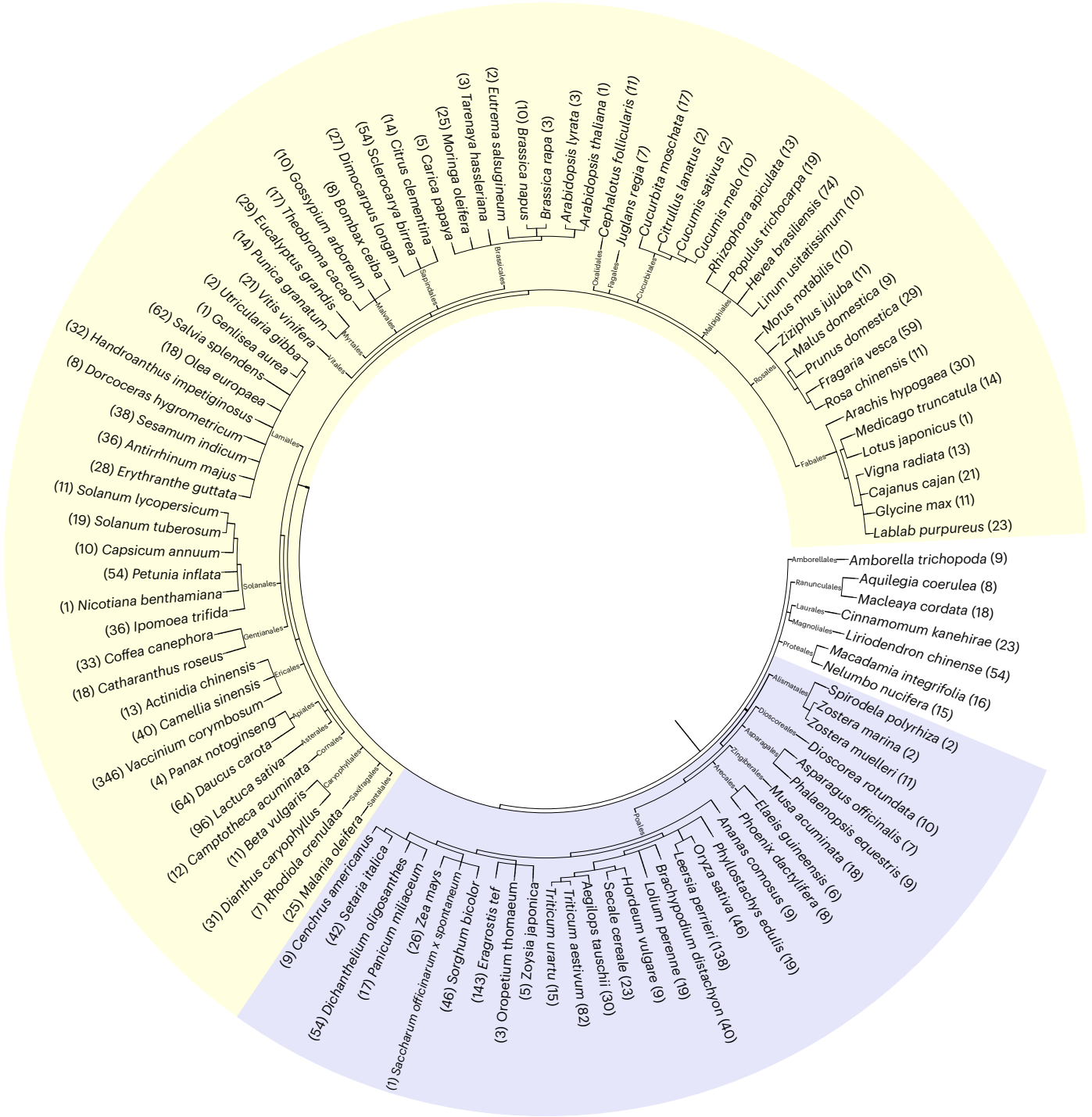


Fig. 1 | A taxonomic tree of 104 plant species was used to generate the TKP Atlas showcasing the number of TKPs. The tree comprises 68 dicot species, 30 monocot species and 7 species from the Magnoliidae clade. The tree is based on the NCBI taxonomy, created using the ete toolkit v3.1.2, showcasing the total number of TKPs in parentheses[48]. This tree illustrates qualitative branching patterns, excluding distance information. Colored species: lilac, monocotyledoneae; yellow, eudicotyledoneae. All of the identified TKPs are available on the Zenodo platform (https://doi.org/10.5281/zenodo.13384335).

presence of kinases likely linked to the membrane, including *L. perrieri* (85 with/55 without transmembrane domain) and *E. tef* (80/63; Supplementary Table 5). The presence of a transmembrane domain alone does not guarantee localization to the plasma membrane, as proteins often require initial signal peptides for this process[34]. We used the SignalP 6.0 tool to identify these signal peptides. Our findings reveal that not every TKP with a transmembrane domain also had detectable signal peptides, suggesting alternative pathways for membrane association, or limitations in detection of signal peptides (Supplementary Table 5). Among the 1,267 TKPs with a transmembrane domain, only 704 (55.5%) harbor a signal peptide, suggesting possible evolution from RLKs (Supplementary Table 5). Indeed, the individual KDs most frequently found in TKPs are also prevalent in RLKs (Fig. 2d).

Surprisingly, over half (55.9%) of identified TKPs contained a non-KD. For our analyses, we did not classify a potential transmembrane domain as an integrated domain. TKP distribution with and without such domains varied across different species. For example, *V. corymbosum* (271 with integrated domain/77 without) and *E. tef* (92/51)

https://doi.org/10.1038/s41588-024-02032-x

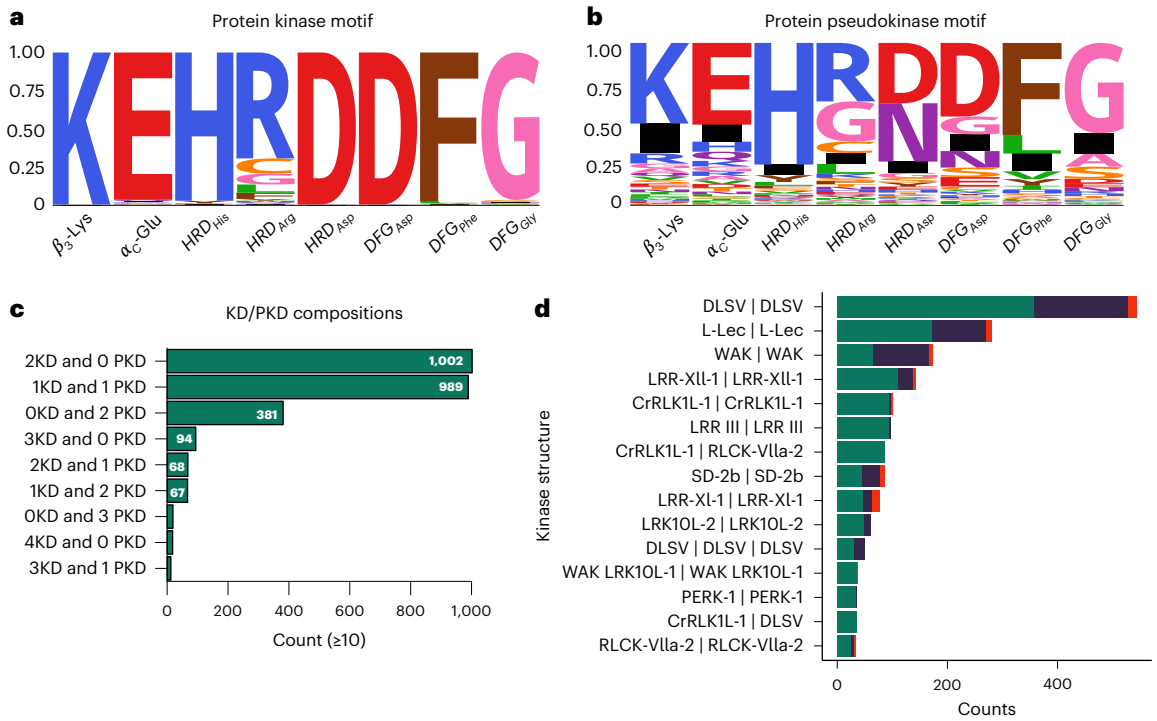


Fig. 2 | Prediction, composition and diversity of kinase and PKDs across 104 plant genomes. a,**b**, Sequence logos demonstrating the residue frequency at key conserved positions in KDs (**a**) and a lack of relative conservation in PKDs (**b**) in each TKP. Amino acids exhibit colors based on their chemical characteristics—positively charged (basic; H, K, R) are represented in blue; negatively charged (acidic; D, E) in red; small, nonpolar, aliphatic (G, A) in pink; aromatic amino acids (F, W, Y) in brown; polar, uncharged amino acids (N, Q) in purple; nonpolar, aliphatic (I, L, M, P, V) in green; and amino acids that contain sulfur or have a hydroxyl group (C, S, T) in orange. **c**, Proportion of TKPs with different combinations of kinase and PKDs across 104 plant genomes. The majority of TKPs comprise two KDs. Kinase/pseudokinase combinations with more than ten members are presented. **d**, Top 15 of most common tandem kinase member counts among TKPs containing two to five kinase or PKDs across 104 plant genomes. Clades are designated by colors—green, eudicotyledoneae; blue, monocotyledoneae; red, magnoliidae. All abbreviations of kinase family names are fully expanded in Supplementary Note.

displayed a higher proportion of proteins with integrated domain (Supplementary Table 5). In contrast, TKPs without integrated domain were more prevalent in *T. aestivum* (11 with/71 without) and *Hevea brasiliensis* (14/60; Extended Data Fig. 3 and Supplementary Table 5).

We identified a total of 127 different classes of integrated domains associated with TKPs. The most prevalent integrated domains were β_lectin (D-mannose binding lectin; PF01453), LRR_8 (LRR VIII, PF13855), and stress–antifungal (or DUF26; PF01657; Fig. 3a). Most integrated domains did not follow a strict positional pattern. However, certain domains such as WD40 (PF00400), NAF (PF03822) and PP2 (PF14299) were found exclusively at the N terminus of the protein (Fig. 3b).

Exploring domain compositions of TKPs with integrated domains, it is notable that the prominent members are proteins containing two KDs from the RLK–Pelle L-Lec and RLK–Pelle LRR III families (Fig. 4a). We also discovered ten TKPs with an integrated heavy-metal-associated (HMA, PS50846) domain, including in the characterized TKP RPG1, which may be involved in the direct recognition of effectors[35,36] (Fig. 4b).

We identified 216 TKPs with integrated domain and presumed cytoplasmic localization due to the lack of transmembrane domain and signal peptides. These kinases contain integrated domains, such as malectin-like, stress–antifungal and legume lectin, that are found in RLKs (Fig. 4b).

Phylogeny of different subfamilies of KDs

To investigate whether TKPs emerged via the duplication of a single gene or the fusion of distinct genes, we performed a phylogenetic analysis of *T. aestivum* tandem kinases that contain two KDs from the same family. We selected this species because all experimentally validated TKPs were found in the Triticeae tribe. We divided the wheat genome into different subgenomes A, B and D for phylogenetic tree construction (Fig. 5). Our findings revealed that most KDs were phylogenetically distant from each other and grouped in separate clades. However, some fusions were between a domain and its most similar domain suggesting potential duplication (Fig. 5). Collectively, this analysis shows that most TKP fusion events in wheat, even within the same kinase family, were not the result of gene duplication. This observation is also supported by the analysis presented in Fig. 6. Thus, duplication may also not explain the presence of other integrated domains.

To investigate the coevolution of different kinase families present within TKPs, we performed multidimensional scaling analyses for the most common families of kinases (L-Lec and DLSV) from all studied species. Our research revealed a clear separation of the tandem KDs from the L-Lec family into the following four distinct clusters: a cluster containing pseudokinases (A), a cluster containing functional kinases (B), and two close-by clusters (C and D; Fig. 6a). Interestingly, we found that pseudokinases from cluster A are fused to KDs from cluster B, suggesting that these two clusters represent one fusion event from which many TKPs evolved. Conversely, kinase and PKD from clusters C and D did not fuse with domains from other clusters, suggesting a tendency for their fusion with very similar domains. The DLSV family contains four clusters where fusion events occurred within each cluster with only two exceptions. A lack of an apparent pattern within each cluster suggests many independent fusion events within very similar domains, but most of them are not gene duplications as the fusions are not between the most similar domains (Fig. 6b). A histogram for two subfamilies was built based on the similarity distance matrix of protein kinases (Extended Data Fig. 4). The histograms demonstrate a diverse distribution of domain distances within a protein for the two families.

https://doi.org/10.1038/s41588-024-02032-x

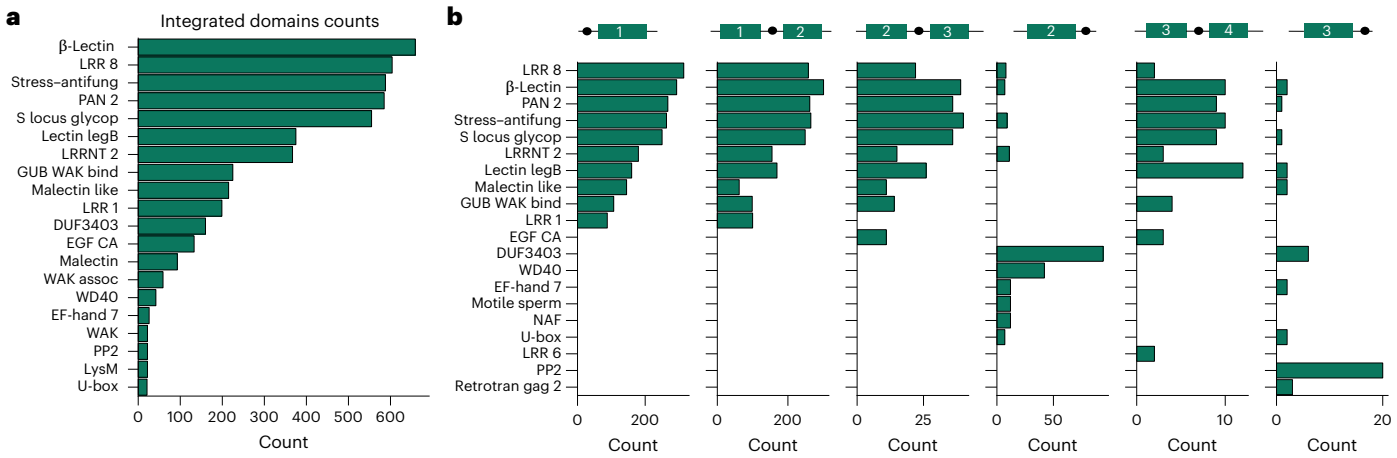


Fig. 3 | Integrated domain prevalence and position relative to protein KDs across 104 plant genomes. a, The number of unique TKPs having a particular integrated domain. Only integrated domains detected 15 or more times are displayed. The transmembrane domain was not counted as an integrated domain. **b**, The location of common integrated domains relative to kinase or PKDs. Top, integrated non-KD (black circle) positioning relative to KDs (green rectangles) enumerated from 1 to 4 serial numbers from C terminus. We displayed only the top 10 most frequent integrated domains per location. Integrations after the third kinase domain were rare.

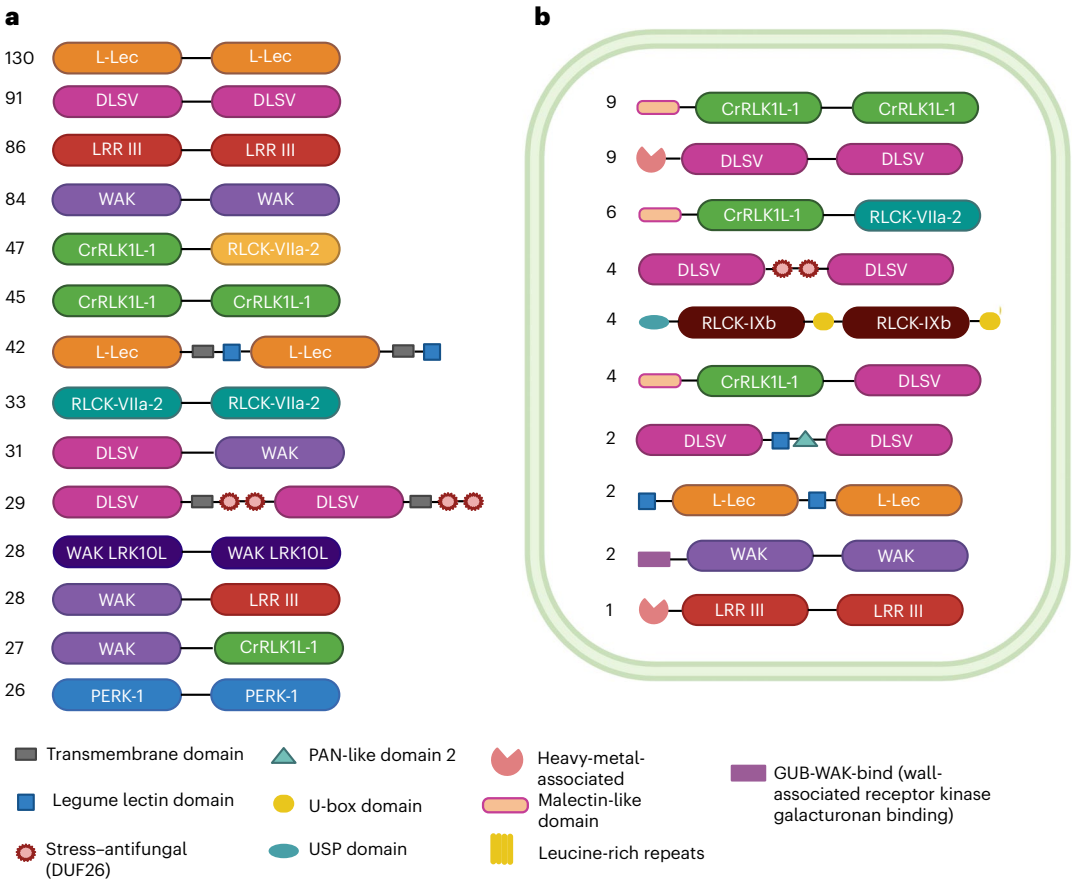


Fig. 4 | Variety of full domain compositions of tandem kinase among 104 plant species. a, Top 14 architectures of TKPs. Numbers indicate the number of TKPs identified with this domain architecture. **b**, Top 10 domain compositions of intracellular TKPs with integrated domains. Numbers indicate the prevalence of each full domain composition. All abbreviations of kinase family names are fully expanded in Supplementary Note. USP, universal stress protein. The figure is created with BioRender.com.

Discussion

TKPs form an important protein family capable of conferring resistance to diverse fungal pathogens. Here we provide an assessment of tandem kinases' distribution across the plant kingdom. Our results demonstrate that the TKP protein family is widespread across the plant kingdom. Recently, two tandem KD-PKDs (TKP7 and TKP8) were found in sugarcane as the candidate genes for *Bru1* brown rust resistance[37]. Wheat blast resistance protein RWT4 can also trigger immunity in rice

https://doi.org/10.1038/s41588-024-02032-x

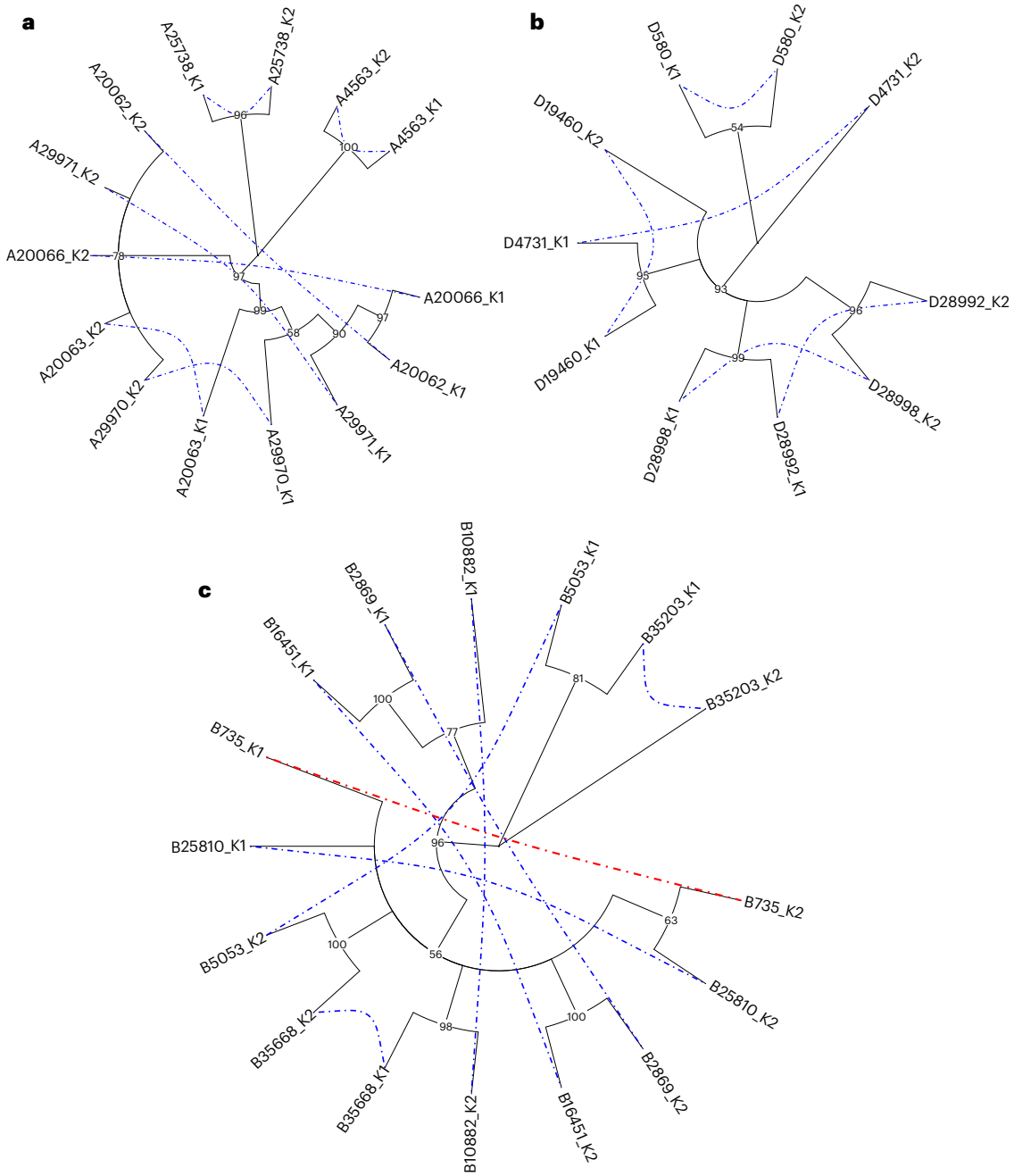


Fig. 5 | Phylogenetic trees of all *T. aestivum* tandem kinases with two domains from DLSV family. a–**c**, Phylogenetic trees are showing tandem kinases from A subgenome (**a**), D subgenome (**b**) and B subgenome (**c**). The red dashed–dotted line shows the connection between the two domains of WTK3 (B735_K1-B735K2). The blue dashed–dotted lines show connections between two domains in the same protein.

protoplasts by perceiving the effector AvrPWT4, indicating TKPs can be transferred between plant genera[22]. Therefore, the current TKP Atlas sheds light on plant immunity, probably far beyond the Triticeae tribe.

The RLK–Pelle kinase family is the plants' most extensive group of RLKs[38]. Within the TKP sequences discovered in this work, more than 95% of the KDs belonged to this family. Our data revealed an almost equal number of TKPs with KD–KD versus KD–PKD architectures. However, predicting unfunctional kinases based only on their sequence is still imperfect and some classified PKDs may possess kinase activity. The kinase activity of RLKs has been well studied, such as the RLK LecRK-IX.2 from *Arabidopsis*[39]. However, the TKP's KD activity was only experimentally proved in the following three cases: RPG1, WTK7-TM and RWT4 (refs. 21,23,40). For example, the two kinases of WTK7-TM were predicted as PKD based on their sequences, but when tested experimentally, they were both shown to be functional[21]. This suggests that different TKPs exhibit diverse kinase activities as should be expected for a protein family that evolved by convergent molecular evolution and provides similar solutions under selection by pathogen stress. Further studies are required to determine the activity and function of the predicted KDs and PKDs.

Most of the experimentally confirmed functional TKPs are cytoplasmic proteins, except for WTK7-TM, which possesses a transmembrane domain shown to be essential for protein function[21]. Furthermore, a quarter of the TKPs in our data had both a transmembrane domain and a signal peptide, indicating membrane association[34]. These accumulated data suggest that at least a portion of TKPs are associated

https://doi.org/10.1038/s41588-024-02032-x

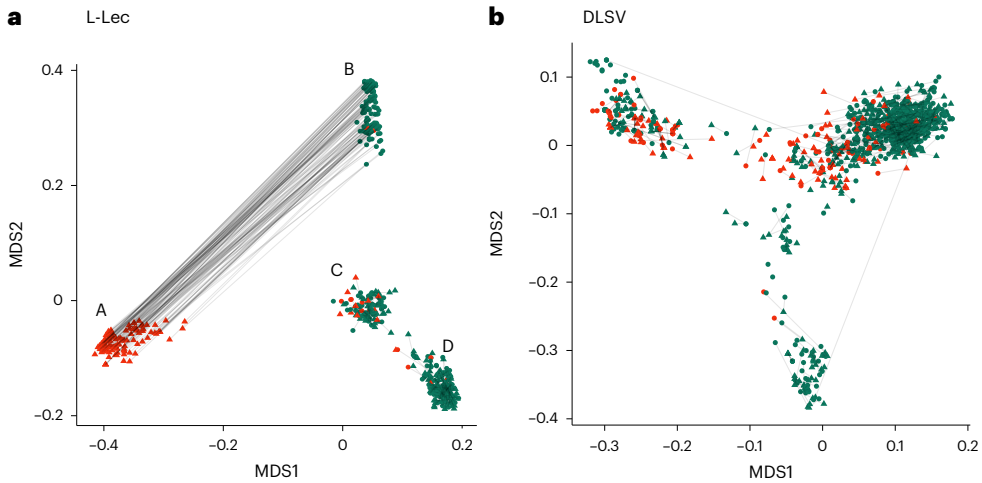


Fig. 6 | Multidimensional scaling analyses of all tandem kinases with two KDs from two RLK–Pelle subfamilies from all studied species. a,**b**, Analyses include KDs from the RLK–Pelle L-Lec (**a**) and DLSV (**b**) subfamilies. A triangle or a circle shows the position of the KD/PKD in the protein—circle, first position; triangle, second position. The color indicates the functionality of the domain—green, kinase; red, pseudokinase. Black lines show connections between two domains in the same protein.

with cell membranes and may detect pathogens in this subcellular location. Taken together, these data provide an important TKP Atlas for the plant community and highlight many facets of TKP evolution and function.

Notably, all ten experimentally identified TKPs contained at least one domain from the RLK–Pelle family[12–22](Supplementary Table 1). Our study suggests that many TKPs with two RLK–Pelle L-Lec domains resulted from a single ancient fusion event between distant family members, persisted across multiple species (monocots and dicots). However, the remainder of TKPs with two RLK–Pelle L-Lec domains arose from numerous independent fusion events among closely related members (Fig. 6). This latter pattern is also typical of the DLSV family. RLKs are a large gene family and can be tandemly clustered, which could facilitate independent fusion events[36]. These independent events may illustrate convergent molecular evolution, particularly if most of the events are associated with plant adaptation to biotic stress mechanisms. Different plant species have independently evolved similar receptor characteristics to adapt to various plant pathogens, as demonstrated for the RLP30 immune receptor, which has a role in immunity against pathogens from two microbial kingdoms[39].

We previously proposed a model for the molecular function of TKPs with a combination of active and inactive domains[41], assuming the PKD serves as a decoy to interact with pathogen components working together with the kinase for signal transduction[41]. However, recently it was shown that RWT4 can specifically bind the AvrPWT4 effector in both the kinase and pseudokinase regions, leading to the transcription of defense genes and inducing cell death[23]. Moreover, TKPs were shown to serve as an unusual class of immune receptors capable of directly interacting with pathogen effectors[23] and activating an NLR helper to trigger immunity[42,43].

Another hypothesis for tandem kinase recognition of the effector occurs via integrated domains. Like in NLR-ID receptors, the integration is the decoy domain, which mimics pathogen effector targets, enabling direct interception of effector proteins[11]. Recently, two TKPs (WTK6-vWA and WTK6b-vWA) were demonstrated to include an integrated domain[20,22]. Mutations in these vWA domains resulted in a loss of resistance, suggesting that the two integrated domains located at the C terminus of these proteins are possibly involved in effector recognition[20]. More than half (56%) of the discovered TKPs contained at least one integrated domain, with the most common being betta lectin, LRR VIII and stress–antifungal. β-Lectin (legume lectin domain, PF00139) is a carbohydrate-binding domain often found in proteins involved in plant defense reactions, such as chitinases, glucanases and thaumatins[25–27]. The function of the LRR VIII domain is poorly understood, but LRR domains are associated with plant immune responses, often found in immune receptors like LRR-RLKs[11]. Another integrated domain identified in NLR proteins is the HMA domain. Multiple NLRs with HMA-integrated domains experimentally verified that the HMA domain directly interacts with the effector[44]. We identified TKPs fused with HMA, including the characterized TKP RPG1. Based on this evidence, it is plausible to propose that integrated domains present in TKPs could also participate in pathogen perception by acting as decoys of pathogen virulence targets.

We identified 637 TKPs with transmembrane domain, signal peptide, and integrated domains (such as stress–antifungal, LysM, β-lectin, etc.), which were also found in another family of plant immune receptors—RLKs[45]. These TKPs probably evolved from RLKs; however, RLKs have one KD, whereas tandem kinases have two or more. This might indicate an evolutionary separation of functions after gene duplication or fusion events, leading to the birth of a new type of receptors in plants[29].

In conclusion, our study provides insights into the diversity of TKPs across the plant kingdom. Our analysis revealed a high degree of variability in the number of these proteins across different species. We also found that many TKPs contain integrated domains, which could affect their functional properties. Our findings highlight the importance of studying TKPs' functional properties in plants and their potential contribution to plant resistance to pathogens. The animal Janus tandem kinases (JAKs) participate in complex regulatory networks governing cell differentiation, tissue regeneration and innate immune responses[46,47]. For example, JAK3 differentially regulates Toll-like receptors-mediated inflammatory cytokine production in innate immune cells. Plants' TKPs may be involved in similar processes, offering another level of functional diversity.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-024-02032-x.

References

1. Savary, S. et al. The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* **3**, 430–439 (2019).
2. Kong, F. & Yang, L. Pathogen-triggered changes in plant development: virulence strategies or host defense mechanism? *Front. Microbiol.* **14**, 1122947 (2023).
3. Boutrot, F. & Zipfel, C. Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. *Annu. Rev. Phytopathol.* **55**, 257–286 (2017).
4. Lu, Y. & Tsuda, K. Intimate association of PRR- and NLR-mediated signaling in plant immunity. *Mol. Plant Microbe Interact.* **34**, 3–14 (2020).
5. Tang, D., Wang, G. & Zhou, J.-M. Receptor kinases in plant–pathogen interactions: more than pattern recognition. *Plant Cell* **29**, 618–637 (2017).
6. Wang, G. et al. A genome-wide functional investigation into the roles of receptor-like proteins in *Arabidopsis*. *Plant Physiol.* **147**, 503–517 (2008).
7. Lolle, S., Stevens, D. & Coaker, G. Plant NLR-triggered immunity: from receptor activation to downstream signaling. *Curr. Opin. Immunol.* **62**, 99–105 (2020).
8. Cesari, S. Multiple strategies for pathogen perception by plant immune receptors. *New Phytol.* **219**, 17–24 (2018).
9. Breit-McNally, C., Laflamme, B., Singh, R. A., Desveaux, D. & Guttman, D. S. ZAR1: guardian of plant kinases. *Front. Plant Sci.* **13**, 981684 (2022).
10. Lüdke, D., Yan, Q., Rohmann, P. F. W. & Wiermer, M. NLR we there yet? Nucleocytoplasmic coordination of NLR-mediated immunity. *New Phytol.* **236**, 24–42 (2022).
11. Dodds, P. N., Chen, J. & Outram, M. A. Pathogen perception and signaling in plant immunity. *Plant Cell* **36**, 1465–1481 (2024).
12. Brueggeman, R. et al. The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. *Proc. Natl Acad. Sci. USA* **99**, 9328–9333 (2002).
13. Klymiuk, V. et al. Cloning of the wheat *Yr15* resistance gene sheds light on the plant tandem kinase–pseudokinase family. *Nat. Commun.* **9**, 3735 (2018).
14. Chen, S. et al. Mapping and characterization of wheat stem rust resistance genes *SrTm5* and *Sr60* from *Triticum monococcum*. *Theor. Appl. Genet.* **131**, 625–635 (2018).
15. Chen, S. et al. Wheat gene *Sr60* encodes a protein with two putative kinase domains that confers resistance to stem rust. *New Phytol.* **225**, 948–959 (2020).
16. Lu, P. et al. A rare gain of function mutation in a wheat tandem kinase confers resistance to powdery mildew. *Nat. Commun.* **11**, 680 (2020).
17. Mago, R. et al. The wheat *Sr50* gene reveals rich diversity at a cereal disease resistance locus. *Nat. Plants* **1**, 15186 (2015).
18. Arora, S. et al. A wheat kinase and immune receptor form host-specificity barriers against the blast fungus. *Nat. Plants* **9**, 385–392 (2023).
19. Yu, G. et al. The wheat stem rust resistance gene *Sr43* encodes an unusual protein kinase. *Nat. Genet.* **55**, 921–926 (2023).
20. Wang, Y. et al. An unusual tandem kinase fusion protein confers leaf rust resistance in wheat. *Nat. Genet.* **55**, 914–920 (2023).
21. Li, M. et al. A membrane associated tandem kinase from wild emmer wheat confers broad-spectrum resistance to powdery mildew. *Nat. Commun.* **15**, 3124 (2024).
22. Liu, W. et al. *Pm57* from *Aegilops searsii* encodes a tandem kinase protein and confers wheat powdery mildew resistance. *Nat. Commun.* **15**, 4796 (2024).
23. Sung, Y.-C. et al. Direct binding of a fungal effector by the wheat RWT4 tandem kinase activates defense. Preprint at *bioRxiv* https://doi.org/10.1101/2024.04.30.591956 (2024).
24. Zhu, K. et al. The grapevine kinome: annotation, classification and expression patterns in developmental processes and stress responses. *Hortic. Res.* **5**, 19 (2018).
25. De Hoff, P. L., Brill, L. M. & Hirsch, A. M. Plant lectins: the ties that bind in root symbiosis and plant defense. *Mol. Genet. Genomics* **282**, 1–15 (2009).
26. Miyakawa, T. et al. A secreted protein with plant-specific cysteine-rich motif functions as a mannose-binding lectin that exhibits antifungal activity. *Plant Physiol.* **166**, 766–778 (2014).
27. Van Holle, S. & Van Damme, E. J. M. Signaling through plant lectins: modulation of plant immunity and beyond. *Biochem. Soc. Trans.* **46**, 217–233 (2018).
28. Bentham, A. R. et al. A molecular roadmap to the plant immune system. *J. Biol. Chem.* **295**, 14916–14935 (2020).
29. Shiu, S.-H. & Bleecker, A. B. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl Acad. Sci. USA* **98**, 10763–10768 (2001).
30. Lehti-Shiu, M. D. & Shiu, S.-H. Diversity, classification and function of the plant protein kinase superfamily. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 2619–2639 (2012).
31. Shakirov, E. V., Chen, J. J.-L. & Shippen, D. E. Plant telomere biology: the green solution to the end-replication problem. *Plant Cell* **34**, 2492–2504 (2022).
32. Li, Q., Jiang, X.-M. & Shao, Z.-Q. Genome-wide analysis of NLR disease resistance genes in an updated reference genome of barley. *Front. Genet.* **12**, 694682 (2021).
33. Kwon, A. et al. Tracing the origin and evolution of pseudokinases across the tree of life. *Sci. Signal.* **12**, eaav3810 (2019).
34. von Heijne, G. The signal peptide. *J. Membr. Biol.* **115**, 195–201 (1990).
35. Cesari, S. et al. New recognition specificity in a plant immune receptor by molecular engineering of its integrated domain. *Nat. Commun.* **13**, 1524 (2022).
36. Białas, A. et al. Two NLR immune receptors acquired high-affinity binding to a fungal effector through convergent evolution of their integrated domain. *eLife* **10**, e66961 (2021).
37. Healey, A. L. et al. The complex polyploid genome architecture of sugarcane. *Nature* **628**, 804–810 (2024).
38. Gish, L. A. & Clark, S. E. The RLK/Pelle family of kinases. *Plant J.* **66**, 117–127 (2011).
39. Xu, N. et al. A plant lectin receptor-like kinase phosphorylates the bacterial effector AvrPtoB to dampen its virulence in *Arabidopsis*. *Mol. Plant* **13**, 1499–1512 (2020).
40. Nirmala, J. et al. Subcellular localization and functions of the barley stem rust resistance receptor-like serine/threonine-specific protein kinase Rpg1. *Proc. Natl Acad. Sci. USA* **103**, 7518–7523 (2006).
41. Klymiuk, V., Coaker, G., Fahima, T. & Pozniak, C. J. Tandem protein kinases emerge as new regulators of plant immunity. *Mol. Plant Microbe Interact.* **34**, 1094–1102 (2021).
42. Chen, R. et al. A wheat tandem kinase sensor activates an NLR helper to trigger immunity. Preprint at *bioRxiv* https://doi.org/10.1101/2024.08.30.610287 (2024).
43. Lu, P. et al. A wheat tandem kinase and NLR pair confers resistance to multiple fungal pathogens. Preprint at *bioRxiv* https://doi.org/10.1101/2024.08.30.610185 (2024).
44. Maqbool, A. et al. Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife* **4**, e08709 (2015).
45. Jose, J., Ghantasala, S. & Roy Choudhury, S. *Arabidopsis* transmembrane receptor-like kinases (RLKs): a bridge between extracellular signal and intracellular regulatory machinery. *Int. J. Mol. Sci.* **21**, 4000 (2020).
46. Wang, H. et al. The role of JAK-3 in regulating TLR-mediated inflammatory cytokine production in innate immune cells. *J. Immunol.* **191**, 1164–1174 (2013).

47. Sarapultsev, A. et al. JAK-STAT signaling in inflammation and stress-related diseases: implications for therapeutic interventions. *Mol. Biomed.* **4**, 40 (2023).
48. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).

Methods

Collection of protein sequence data across the plant kingdom

Annotated protein sequences of 104 plant species were obtained from several databases—NCBI, Origin, Giga, JDI, Dryad, PLAZA, GDR and Ensembl. We included plants from three different clades in the analysis—Magnoliidae (7 species), Monocotyledoneae (30 species) and Eudicotyledoneae (68 species; Supplementary Table 2 and Fig. 1).

Identification of tandem kinases among all proteins

Almost 5 million protein sequences (4,707,304) of 104 seed plant species were aligned against the protein KD's profile (PS50011) by ps_scan software from ProSITE (https://prosite.expasy.org/). Hits with less than 150 amino acids were discarded. Putative TKPs were defined as protein sequences with two or more protein KDs arranged in tandem; all proteins with a single KD were discarded. Longer alternative transcripts were chosen in case of multiple annotated transcripts for one gene.

The annotation of the kinase units of TKPs

To annotate protein KDs at the subfamily level, we extracted sequences of all KDs from the iTAK database, which was built on the protein kinase classification in ref. 30. Next, we used these sequences to create a profile database using HMMER (v3.1b2). Then, we used hmmscan from HMMER with our KDs against this database and annotated it based on the most substantial entry. Currently, there is no stable definition or defined profile for a protein PKD. Therefore, we applied an empirical conservation-based approach, targeting catalytically essential residues, namely amino acid substitutions that are critical for phosphorylation activity at one of the three conserved sites of the 'catalytic triad'—ATP-binding β3 lysine (K), the catalytic aspartate within the catalytic loop (HRD), and the metal-binding aspartate of the activation loop (DFG)[33]. If domains contained mutations or deletions in any of the three amino acids, we classified these domains as pseudokinases. Multiple sequence alignments (MSA) of all KDs were performed by *mafft* (v7.130). We pulled catalytic triad residues out of the MSA and used them to classify sequences as KD/PKD. This was based on whether these residues were present or substituted in one or more of the extracted positions. Using this definition, we classified domains as likely functional kinases or pseudokinases.

Identification and classification of integrated domains fused to tandem kinases

We used HMMER with all available Pfam-A profiles to annotate integrated domains fused to tandem kinases. We used the ProSITE tool for the annotation of the HMA (PS50846) domain due to its small size, which rendered it challenging to identify using HMMER. We also predicted transmembrane domains with Phobius and TMHH web services, accepting domains identified by both tools. Signal peptides in the sequences were identified using SignalP 6.0 (ref. 49). Cellular localization of proteins carrying signal peptides was predicted using DeepLoc 2.0 (ref. 50). All subsequent analyses were performed in the R environment for statistical computing (https://www.r-project.org/).

Phylogenetic and multidimensional scaling analyses

We employed two filters in our dataset to elucidate the complex phylogenetic history of tandem kinases. Specifically, we only included TKPs possessing two KDs and those exhibiting equivalent annotation at the subfamily level for both KDs. We extracted KDs from the filtered sequences and grouped them by species and subfamily. We performed multiple sequence alignment with mafft (v7.130) and dropped out all outlier sequences by Sequence Bouncer with a 10% gap cutoff. We constructed phylogenetic trees for tandem kinases with two domains from the RLK–Pelle DLSV subfamily from *T. aestivum*. We constructed 1,000 bootstrap replicate trees using RAxML to evaluate the robustness of the phylogenetic relationships. Finally, we calculated pairwise distances for sequences from several considered subfamilies by dist.alignment function from seqinr (version 4.2-16) R package and performed multidimensional scaling analyses with R base function cmdscale[51,52]. Then, we constructed a frequency histogram depicting the distribution of similarity scores.

Statistical tests

We tested two hypotheses related to our results—first, whether the count of TKPs is correlated with species genome size, and second, whether the sequencing technology used to create the assembly is affecting the TKP counts. For the first hypothesis, we used the Spearman correlation test because our data does not follow a normal distribution and lacks linearity (Supplementary Table 2).

To examine whether the sequencing technology used to create the assembly is affecting the number of detected TKPs, we divided our dataset into the following three groups: (1) genomes sequenced using short-read-based technology, (2) long-read-based technology and (3) a hybrid of these technologies. We then performed an ANOVA test to compare the mean number of TKPs across these groups (Supplementary Table 2).

Reporting summary

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

Data availability

We have published a fasta file with all found TKPs on the Zenodo platform with a link: https://doi.org/10.5281/zenodo.13384335 (ref. 53).

Code availability

We have published our code on the Zenodo platform with a link: https://doi.org/10.5281/zenodo.11118417 (ref. 54).

References

49. Teufel, F. et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**, 1023–1025 (2022).
50. Thumuluri, V., Almagro Armenteros, J. J., Johansen, A. R., Nielsen, H. & Winther, O. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res.* **50**, W228–W234 (2022).
51. Dunn, C. D. SequenceBouncer: a method to remove outlier entries from a multiple sequence alignment. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.24.395459 (2021).
52. Bastolla, U., Porto, M., Roman, H. E. & Vendruscolo, M. (eds) *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*, pp. 207–232 (Springer, 2007).
53. Fatiukha, A., Potapenko, E. & Reveguk, T. All tandem kinase proteins from 104 plant genomes. *Zenodo* https://doi.org/10.5281/zenodo.13384335 (2024).
54. Fatiukha, A. R_scrips_for_TKP_atlas. *Zenodo* https://doi.org/10.5281/zenodo.11118417 (2020).

Acknowledgements

The authors thank O. Borzov, A. Korol, T. Krugman and L. Govta from the Institute of Evolution at the University of Haifa for their professional and moral support. T.F. was supported by the United States-Israel Binational Science Foundation (2019654), the US-Israel Binational Agricultural Research and Development Fund (US-5191-19C and US-5515-22C) and the Israel Science Foundation (grants 1366/18 and 2342/18). A.F. was supported by the EU COST INDEPH (CA16212). G.C. and T.F. were supported by the United States National Science Foundation (1937855) and the United States Department of Agriculture (2020-67013-32577). G.C. was supported by a grant from the National Institutes of Health (R35GM136402).

Author contributions

T.F. conceived the overall research concept. A.F. and T.W. developed the pipeline for searching the TKPs, which A.F. used with the genomes

of 104 plants. T.F., A.F., V.K., H.S., Y.L., G.C., T.R., E.P. and C.P. designed the experimental approach. E.P., T.R. and I.R. analyzed the TKPs data. T.R. and H.S. conducted the phylogenetic and multidimensional scaling analyses. V.K. and Y.L. contributed to discussions and provided valuable advice on interpreting the research findings. T.R., T.F. and G.C. wrote the initial version of the manuscript. All authors contributed to subsequent versions and have read and approved the manuscript.

Competing interests
The authors declare no competing interests.

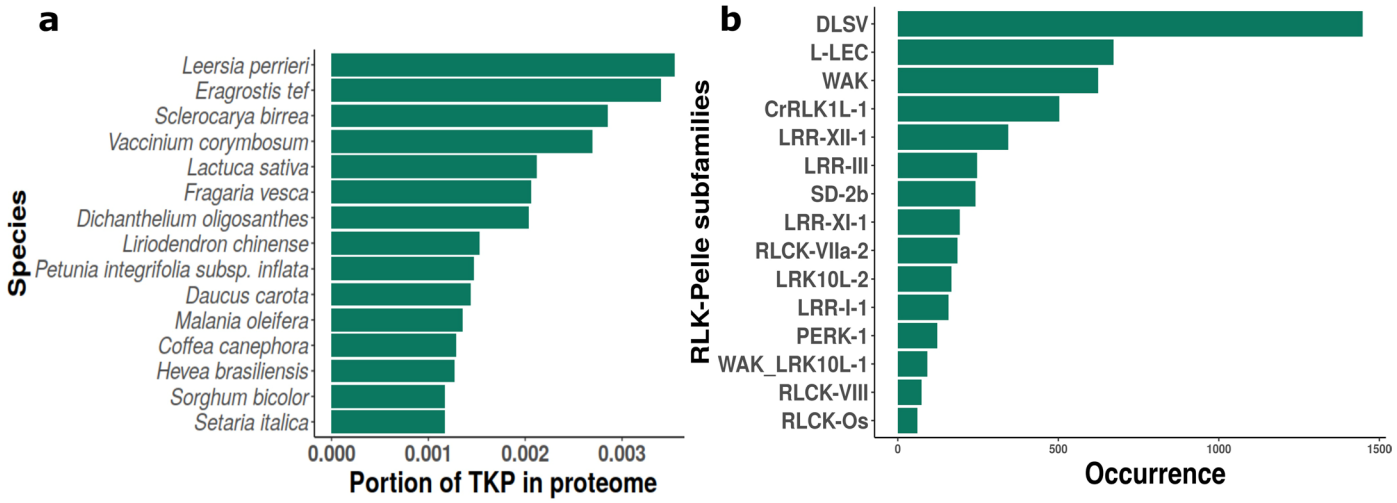
Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41588-024-02032-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-024-02032-x.

Correspondence and requests for materials should be addressed to Gitta Coaker or Tzion Fahima.

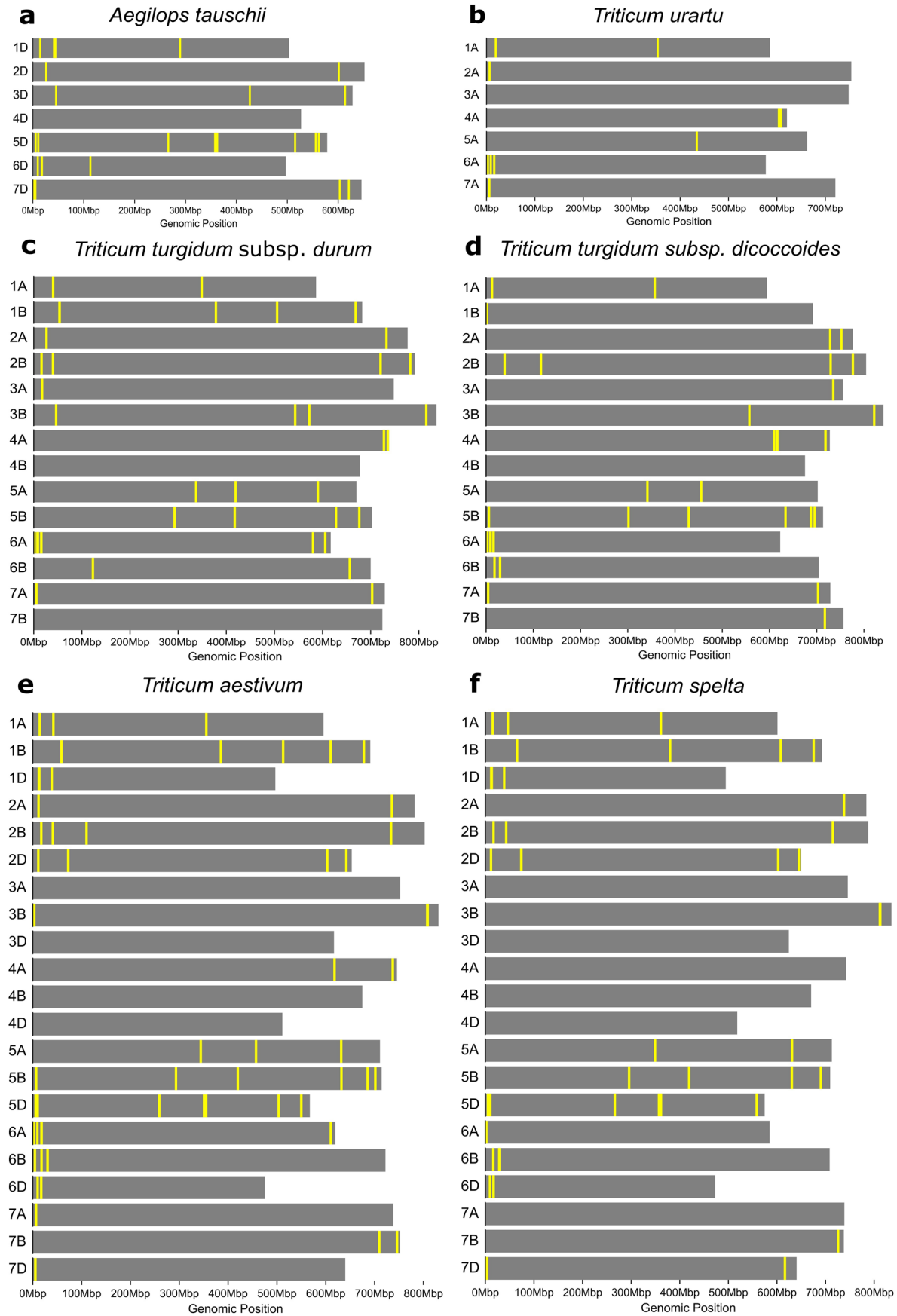
Peer review information *Nature Genetics* thanks Manuel Spannagl and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41588-024-02032-x



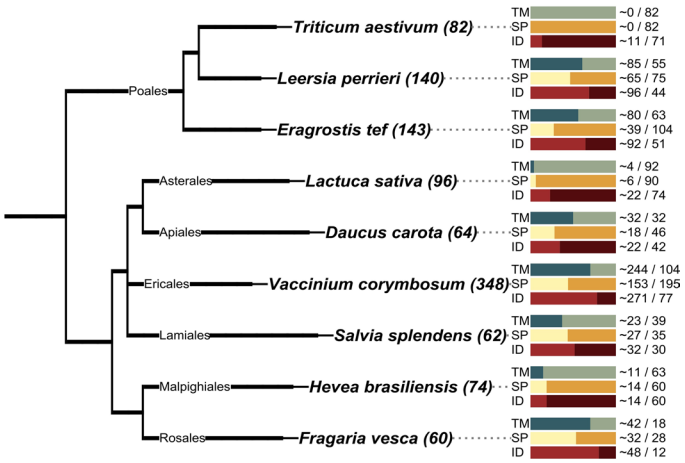
Extended Data Fig. 1 | Domain diversity and proportion of TKPs identified from predicted proteomes. a, The proportion of TKPs compared to the total proteome size. **b**, TKPs containing two to five kinase domains and their family membership.

https://doi.org/10.1038/s41588-024-02032-x



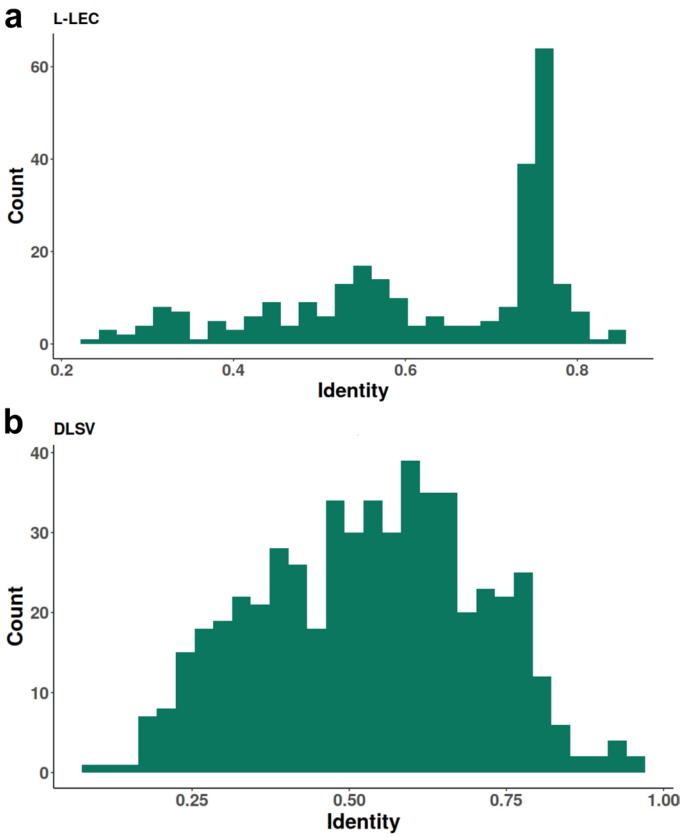
Extended Data Fig. 2 | Genomic position of TKPs on chromosomes of cereal species. a, *Aegilops tauschii* 2n = 2x = 14, DD; **b**, *Triticum urartu* 2n = 2x = 14, AA; **c**, *Triticum turgidum* subsp. *durum* 2n = 4x = 28, AABB; **d**, *Triticum turgidum* subsp. *dicoccoides* 2n = 4x = 28, AABB; **e**, *Triticum aestivum* 2n = 6x = 42, AABBDD; **f**, *Triticum spelta* 2n = 6x = 42, AABBDD. Yellow lines indicate the TKPs positions.

https://doi.org/10.1038/s41588-024-02032-x



Extended Data Fig. 3 | A tree with the nine species having the most abundant number of TKPs. Bars demonstrate the presence/absence of a transmembrane region (TM), a signal peptide (SP) and at least one integrated (nonkinase) domain (ID). The tree represents taxonomy, created using the ete toolkit v3.1.2, showcasing total TKP counts as a bar chart. This tree illustrates qualitative branching patterns, excluding distance information.

https://doi.org/10.1038/s41588-024-02032-x



Extended Data Fig. 4 | Distance histogram calculated on domain sequences with matching family annotations. a, For TKPs with two domains from the RLK–Pelle Lec family. **b**, For TKPs with two domains from RLK–Pelle DLSV family. Pairwise distances were calculated for sequences from two subfamilies by dist.alignment function from seqinr (version 4.2-16) R package.

nature portfolio

Corresponding author(s): Tzion Fahima, Gitta Coaker

Last updated by author(s): Oct 8, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We did not use any software or code for data collection.
Data analysis	For data analysis, we used publicly available software and custom-developed code. Specifically, we used PS_SCAN software from ProSITE, HMMER (v3.1b2), MAFFT (v7.130), SignalP 6.0, DeepLoc 2.0, and the dist.alignment function from the seqinr (version 4.2-16) R package. Our code is published on Zenodo (https://doi.org/10.5281/zenodo.111184178).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

In this study, we used annotated protein sequences from 104 plant species, which were obtained from several databases: NCBI, OrthoDB, GigaDB, JGI, Dryad,

PLAZA, GDR, and Ensemble plants. All references to the 104 genome studies are listed in Supplementary Table 2. Additionally, we used the iTAK database for annotating protein kinase domains and all available Pfam-A profiles to annotate integrated domains.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	Annotated protein sequences of 104 plant species were obtained from several databases: ncbi, origin, giga, JDI, dryad, plaza, GDR, and Ensemble. We included in the analysis plants from three different clades: Magnoliidae (7 species), Monocotyledoneae (30 species), and Eudicotyledoneae (68 species).
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	All data from the selected annotated genomes are presented in the supplementary material.
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*