

## BIOCHEMISTRY

# Tracing the origin and evolution of pseudokinases across the tree of life

Annie Kwon<sup>1,2</sup>, Steven Scott<sup>2,3</sup>, Rahil Taujale<sup>1,2</sup>, Wayland Yeung<sup>1,2</sup>, Krys J. Kochut<sup>4</sup>,  
 Patrick A. Eyers<sup>5</sup>, Natarajan Kannan<sup>1,2\*</sup>

Copyright © 2019  
 The Authors, some  
 rights reserved;  
 exclusive licensee  
 American Association  
 for the Advancement  
 of Science. No claim  
 to original U.S.  
 Government Works

**Protein phosphorylation by eukaryotic protein kinases (ePKs) represents a fundamental mechanism of cell signaling in all organisms. In model vertebrates, ~10% of ePKs are classified as pseudokinases, which have amino acid changes within the catalytic machinery of the kinase domain that distinguish them from their canonical kinase counterparts. However, pseudokinases still regulate various signaling pathways, usually doing so in the absence of their own catalytic output. To investigate the prevalence, evolutionary relationships, and biological diversity of these pseudoenzymes, we performed a comprehensive analysis of putative pseudokinase sequences in available eukaryotic, bacterial, and archaeal proteomes. We found that pseudokinases are present across all domains of life, and we classified nearly 30,000 eukaryotic, 1500 bacterial, and 20 archaeal pseudokinase sequences into 86 pseudokinase families, including ~30 families that were previously unknown. We uncovered a rich variety of pseudokinases with notable expansions not only in animals but also in plants, fungi, and bacteria, where pseudokinases have previously received cursory attention. These expansions are accompanied by domain shuffling, which suggests roles for pseudokinases in plant innate immunity, plant-fungal interactions, and bacterial signaling. Mechanistically, the ancestral kinase fold has diverged in many distinct ways through the enrichment of unique sequence motifs to generate new families of pseudokinases in which the kinase domain is repurposed for noncanonical nucleotide binding or to stabilize unique, inactive kinase conformations. We further provide a collection of annotated pseudokinase sequences in the Protein Kinase Ontology (ProKinO) as a new mineable resource for the signaling community.**

## INTRODUCTION

Protein phosphorylation catalyzed by eukaryotic protein kinases (ePKs) controls multiple aspects of prokaryotic- and eukaryotic-based cell signaling (1, 2), and its dysregulation contributes to many major diseases. The conserved architecture of the ePK domain is very well understood from both structural (3–5) and biochemical (6–8) perspectives, and the versatility of the kinase fold has been exploited many times during evolution to impart mechanistic control over diverse cell signaling processes (9, 10). A vast amount of genomic and proteomic datasets can now be mined to map the evolution of kinases and their associated signaling pathways across multiple species (11–17). In this context, some 10% of model vertebrate ePKs contain amino acid changes at specific positions that are predicted to lead to catalytic inactivation, which led to the coining of the term “pseudokinase” (5, 15, 18–21). A number of well-studied pseudokinases are thought to play central roles in signaling despite impaired catalytic function (22–26), for example, through allosteric modulation of other active kinases or the transduction of cellular signals via dynamic scaffolding functions (9, 19, 21, 27–30). However, whether pseudokinases have evolved to control fundamental aspects of signaling across all organisms has never been scrutinized in depth, and much remains to be understood about the origin of pseudokinases and how they became embedded in signaling networks during prokaryotic and eukaryotic evolution.

Protein pseudokinases represent the best understood members of the growing classes of pseudoenzymes, which include pseudophosphatases (31) and pseudoproteases (32), both of which are also predicted to have lost canonical catalytic function but nonetheless perform critical non-enzymatic roles (9, 20, 33, 34). By definition, pseudokinases lack canonical phosphotransferase activity, and they can be predicted bioinformatically by identifying sequences that lack at least one key amino acid normally required for metal and adenosine triphosphate (ATP) binding and for catalysis (3, 7, 8, 18–20). Prominent catalytic motifs include the “catalytic triad” residues, composed of the ATP-binding β3-lysine, the catalytic aspartate within the catalytic loop His-Arg-Asp-X-X-X-Gln (HRDXXXXN) motif, and the metal-binding aspartate of the activation loop Asp-Phe-Gly (DFG) motif. Some examples of human pseudokinases with variations at these catalytic triad residues are summarized in Table 1. Loss of these canonical residues does not always abolish nucleotide binding or phosphoryl transfer, and in some cases, residual kinase activity or ATP binding may fulfill a unique functional role. However, we still define these catalytically competent proteins as pseudokinases, in recognition of their noncanonical amino acid composition. For example, in the human epidermal growth factor receptor (EGFR)-related receptor pseudokinase HER3, where the catalytic triad is conserved except for the substitution of the catalytic HRD-aspartate for asparagine, low amounts of catalytic activity support HER3 trans-autophosphorylation *in vitro*, although this vestigial activity is insufficient for phosphorylation of exogenous substrates in cells (22, 35, 36). In other cases, degenerated catalytic residues can be compensated by similar amino acids found elsewhere in the active site to rescue catalytic function. This is best illustrated by the with-no-lysine kinases (WNKs), which lack the canonical β3-lysine but maintain ATP binding and catalytic activity via a conserved compensatory lysine in the glycine-rich loop (37–39). Less predictably, pseudokinases contain coevolving amino acids

<sup>1</sup>Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA. <sup>2</sup>Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA. <sup>3</sup>Department of Genetics, University of Georgia, Athens, GA 30602, USA. <sup>4</sup>Department of Computer Science, University of Georgia, Athens, GA 30602, USA. <sup>5</sup>Department of Biochemistry, Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK.

\*Corresponding author. Email: nkannan@uga.edu

AQ16  
AQ65AQ66

**Table 1. Examples of human pseudokinases.** Degraded catalytic triad residues and the amino acids that replace them in each pseudokinase are noted. ILK, integrin-linked kinase; VRK, vaccinia-related kinase.

Human pseudokinase	Degraded catalytic residue(s)	Observed residue(s)
KSR1, KSR2	K	R
WNK1, WNK2, WNK3, WNK4	K	C
HER3	HRD-D	N
JAK1 (domain2)	HRD-D	N
JAK2 (domain2)	HRD-D	N
ILK	HRD-D	A
TRIB3	DFG-D	N
TRIB2	DFG-D	S
CASK	DFG-D	G
GCN2 (domain2)	K, HRD-D	Y, V
ULK4	K, DFG-D	L, N
VRK3	HRD-D, DFG-D	N, G
MLKL	HRD-D, DFG-D	K, G
STRADB (STLK6)	HRD-D, DFG-D	S, G
EphB6	K, HRD-D, DFG-D	Q, S, R
SCYL1, SCYL2, SCYL3	K, HRD-D, DFG-D	F, N, G
NRBP1, NRBP2	K, HRD-D, DFG-D	N, N, S

that are far removed from the active site and contribute critical non-catalytic signaling roles, as described recently for the Tribbles (TRIB) family of pseudokinases, where a coevolved C-terminal tail docking site in the pseudokinase domain negatively regulates binding of the E3 ubiquitin–protein ligase COP1 (26, 40–42). Last, amino acid shuffling offers new biochemical opportunities as described recently for the atypical selenoprotein-O (SelO) pseudokinase in which marked active site variations facilitate “inverted” ATP binding to support protein AMPylation instead of phosphorylation (43, 44).

About 50 protein pseudokinases are encoded by the human genome, nearly all of which are also found in rodents, suggesting conserved vertebrate-wide signaling roles (5, 15, 18). Half of vertebrate pseudokinases also have clear orthologs in well-characterized genetic model organisms, including flies and worms, supporting the assumption that pseudokinases are part of ancient genetic lineages, rather than extraneous remnants of evolutionary “experiments” (13). Several pseudokinases have been analyzed in depth in human cells, including HER3 (23, 35, 45, 46), the RAF/mitogen-activated protein kinase (MEK) modulators kinase suppressor of Ras-1 and Ras-2 (KSR1/2) (24), the Janus tyrosine kinases (JAKs) (25), which contain a disease-associated pseudokinase domain positioned adjacent to an active kinase domain, and the TRIB family of pseudokinases (26, 40, 47). However, more than half of the predicted human pseudokinome remains understudied at the molecular level, despite clear evidence for expression in cells. Pseudokinase-based signaling has also been described in simple model organisms (48), such as in the small genome of the intestinal parasite *Giardia lamblia* (16) and in commercially important plants (49, 50, 51, 52, 53). However, the origin and evolution of pseudokinases across the tree of life has not been explored in any depth.

In this resource, we present a systematic identification of predicted pseudokinase sequences, ranging from archaea and bacteria

to simple eukaryotes, fungi, plants, and vertebrates. On the basis of the well-understood catalytic machinery in canonical ePKs (6), we find that ePK-like pseudokinase sequences can even be detected in some archaeal and bacterial proteomes, although they are much rarer than in eukaryotic proteomes, where they appear to be ubiquitous. Corroborating previous kinome studies, we also find that the number of pseudokinases remains relatively constant among vertebrates and correlates with the relative size of the kinome. Our broad analysis permits us to establish that ~10% of ePK members should be classified as pseudokinases in swathes of vertebrate animal species. In several phyla, specific pseudokinase expansions linked to lifestyle are observed within the different kinase families, whose shared sequence signatures and domain structures permit specific functions to be deciphered. In particular, we note the expansions of interleukin-1 receptor–associated kinase (IRAK)–like pseudokinases in plants, increases of tyrosine kinase–like (TKL) pseudokinases in fungi, and a diversified family of protein kinase B (PknB) pseudokinases in bacteria. Most pseudokinases exhibit lineage-specific sequence variations that might facilitate previously unknown modes of ATP binding, unusual catalytic outputs, and/or allosteric coupling between distal protein binding and regulatory sites. Hence, pseudokinases cannot be remnants of evolution but must instead operate as fundamental, and function-specific, signaling proteins across organisms covering some 4 billion years of evolution. Our analysis includes a minable, comprehensive classification of pseudokinase sequences from diverse organisms, providing a conceptual starting point for future hypothesis-driven characterization of pseudokinase signaling from bacteria to humans.

## RESULTS

### Identification of pseudokinomes across the domains of life

To detect the prevalence of pseudokinases across the domains of life, we used curated multiple sequence alignment profiles of 592 protein kinase families (52, 54–58) to scan all available eukaryotic, bacterial, and archaeal proteomes in the UniProt reference proteome database (10,092 proteomes in release 2018\_9) (59). Aligned sequences that lacked one or more of the canonical catalytic triad residues, namely, the  $\beta$ -3-lysine residue, the HRD-aspartate, or the DFG-aspartate, were classified as pseudokinase sequences. Such pseudokinase sequences with noncanonical residues at the three catalytic triad positions were detected in 100% of eukaryotic proteomes analyzed, whereas only 5.8 and 2.5% of bacterial and archaeal proteomes, respectively, contained putative pseudokinase sequences (Table 2). The prevalence of kinases and pseudokinases across 93 diverse representative species throughout the domains of life is summarized in Figs. 1 and 2. Consistent with previous studies, we identified 55 pseudokinases in the human kinome, including pseudopodium-enriched atypical kinase 3 (PEAK3), which was recently reported as a pseudokinase sharing similarity to the pragmin and PEAK1/Sgk269 pseudokinases (Fig. 1) (30, 60). We also identified a previously unidentified putative human pseudokinase (A0A1B0GUL7) that is homologous to the dual-specificity MEK2 and has a pseudokinase ortholog in chimpanzee (*Pan troglodytes*). Other previously studied pseudokinases such as transformation/transcription domain-associated protein (19, 61), SelO (43, 44), and family with sequence similarity 20 A (Fam20A) (29) are also considered pseudokinases; however, these atypical kinases and other small-molecule kinases such as aminoglycoside kinases and lipid kinases were not considered in our analysis

**Table 2. Protein kinase and pseudokinase sequences across archaeal, bacterial, and eukaryotic proteomes.** Detection of protein kinase and pseudokinase sequences across archaeal, bacterial, and eukaryotic proteomes available from the UniProt database (59). Protein kinase sequences were identified and aligned from each reference proteome using diverse sequence profiles of ePKs, and pseudokinase sequences were identified on the basis of their lack of at least one residue in the catalytic triad.

	Archaea	Bacteria	Eukaryota
<b>Total no. of proteomes</b>	441	8818	833
<b>Proteomes with kinases</b>	149 (33.8%)	3523 (40.0%)	833 (100%)
<b>Proteomes with pseudokinases</b>	11 (2.49%)	508 (5.76%)	819 (98.3%)
<b>Total no. of pseudokinase sequences</b>	20	1386	30,049

because they are markedly divergent from ePKs and cannot be reliably placed within the ePK evolutionary framework (further described in Materials and Methods).

Pseudokinase complements of kinomes (hereafter referred to as pseudokinomes) are nearly always proportional to the size of the kinome in vertebrate species (Fig. 1), which is consistent with previous estimates that pseudokinases generally account for ~10% of kinome content (18). However, both kinome and pseudokinome sizes are much more diverse across other (nonvertebrate) eukaryotic clades. For example, pseudokinase sequences account for between 8 and 17% of plant kinomes, which are often markedly expanded in size when compared to metazoan kinomes. Moreover, a mycorrhizal fungal species, *Rhizophagus irregularis*, and two protist species, *Paramecium tetraurelia* and *Tetrahymena thermophila*, have substantially expanded kinomes relative to other fungi and protists, and the pseudokinomes of *R. irregularis* and *T. thermophila* are also markedly expanded, comprising 32 and 25% of each kinome, respectively. We also note a remarkable expansion of pseudokinases in eukaryotic pathogens, including *Plasmodium falciparum* and *G. lamblia*, which have relatively small kinomes (16, 62) but contain highly expanded pseudokinomes that account for more than a half of their kinomes (Fig. 1).

Also unprecedented was the varied detection of pseudokinases across diverse bacterial phyla with high sequence similarity to PknB kinases (Fig. 2). Bacterial kinomes and pseudokinomes exhibit a large amount of diversity in size, particularly when compared to those in eukaryotes. For example, we note the large expansion of pseudokinases in *Streptomyces coelicolor*, which has 31 protein kinases and 5 pseudokinases, whereas the proteomes of *Shigella flexneri* and *Escherichia coli* like most bacterial proteomes lack pseudokinases, containing only one protein kinase sequence each. In contrast, the proteomes of some bacterial species, including *Treponema denticola*, do not contain any detectable protein kinase sequences. Nonetheless, pseudokinases were detected in at least one species for every bacterial phylum that we examined, except in *Chlamydiae*, suggesting that pseudokinases are not confined to any specific bacterial classifications such as Gram-negative or Gram-positive bacteria but rather are found across diverse bacterial phyla.

A small number of ePK sequences were also detected in archaea (Fig. 2 and table S1), where 11 archaeal proteomes contained putative pseudokinase sequences. In particular, although most of these

archaeal proteomes contained only one or two pseudokinase sequences, seven of the eight ePK sequences detected in *Halorientalis regularis* were identified as pseudokinases. Alignments for the kinomes and pseudokinomes of every organism represented in the main text (Figs. 1 and 2) are provided in the Supplementary Materials (data file S1). Also in the Supplementary Materials, the number of ePK and pseudokinase sequences detected in the 10,092 proteomes currently available in the UniProt reference proteome database is additionally provided (table S1).

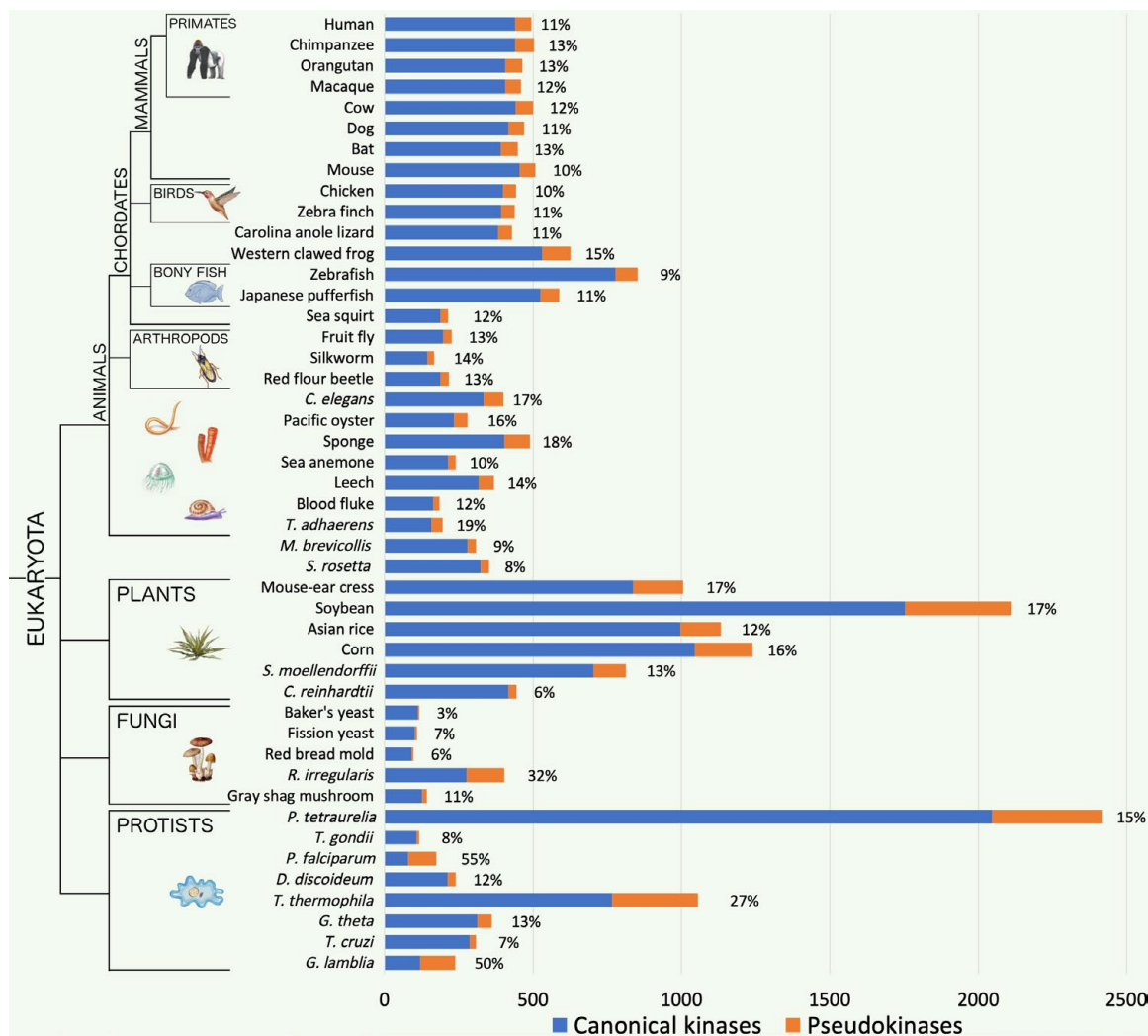
### Classification of pseudokinase sequences

We next classified pseudokinase sequences into evolutionarily related clusters using an optimal multiple-category Bayesian Partitioning with Pattern Selection (omcBPPS) algorithm, which classifies sequences based on patterns of amino acid conservation and variation in a large multiple sequence alignment (see Materials and Methods) (63, 64). Because of the large number of pseudokinase sequences analyzed, we first classified a diverse representative set of 26,273 pseudokinase sequences (see Materials and Methods for details). Some of the well-known metazoan pseudokinase families such as the JAK and TRIB families were found to fall into distinct pseudokinase sequence clusters (26). The identified pseudokinase clusters were incorporated within an existing evolutionary hierarchy of kinase groups and families (see Materials and Methods) (18), and a resulting hierarchy of 592 sequence profiles containing both canonical kinase families and pseudokinase families was used to detect, classify, and align all pseudokinase sequences from the UniProt reference proteome database.

Overall, we detected pseudokinase families across all major kinase groups (Fig. 3). Whereas some kinase groups defined by Manning and colleagues (18), such as the STE and AGC groups, have relatively few pseudokinases, the TKL group is highly enriched with pseudokinases. This increase is caused by the diversification of TKL pseudokinase sequences in various plant and fungal species, where the general expansion of canonical TKL kinases has previously been noted (52, 65). We also identify species-specific divergence of pseudokinase sequences, such as the divergence of metazoan-specific poly(A)-nuclease deadenylation complex subunit 3 (PAN3) from other eukaryotic PAN3, and the diversification of plant WNK kinases from the classical chordate WNK kinases. In contrast, some pseudokinase clusters comprise orthologs found across diverse taxonomic groups, such as Haspin, which is found across diverse eukaryotes, and the TRIB family of pseudokinases, which is found across diverse metazoa, where they function as regulators of protein ubiquitylation (26). Detailed taxonomic annotations for each pseudokinase family are provided in the Supplementary Materials (table S2).

Using our integrated evolutionary hierarchy of pseudokinase clusters with previously established kinase groups and families, we can now identify some canonical kinase sequences conserving the catalytic triad residues that classify into pseudokinase families. For example, four sequences containing the canonical HRD motif classified with the HER3 pseudokinases, which typically contain a conserved HRN motif that severely blunts catalysis. These sequences were identified to be HER3 orthologs in four separate rodent species (*Mus musculus*, *Rattus norvegicus*, *Cricetulus griseus*, and *Mesocricetus auratus*) (66). Similarly, whereas orthologs of the KSR family can be detected across diverse metazoan species, only chordate KSRs are classified as predicted pseudokinases due to the replacement of the



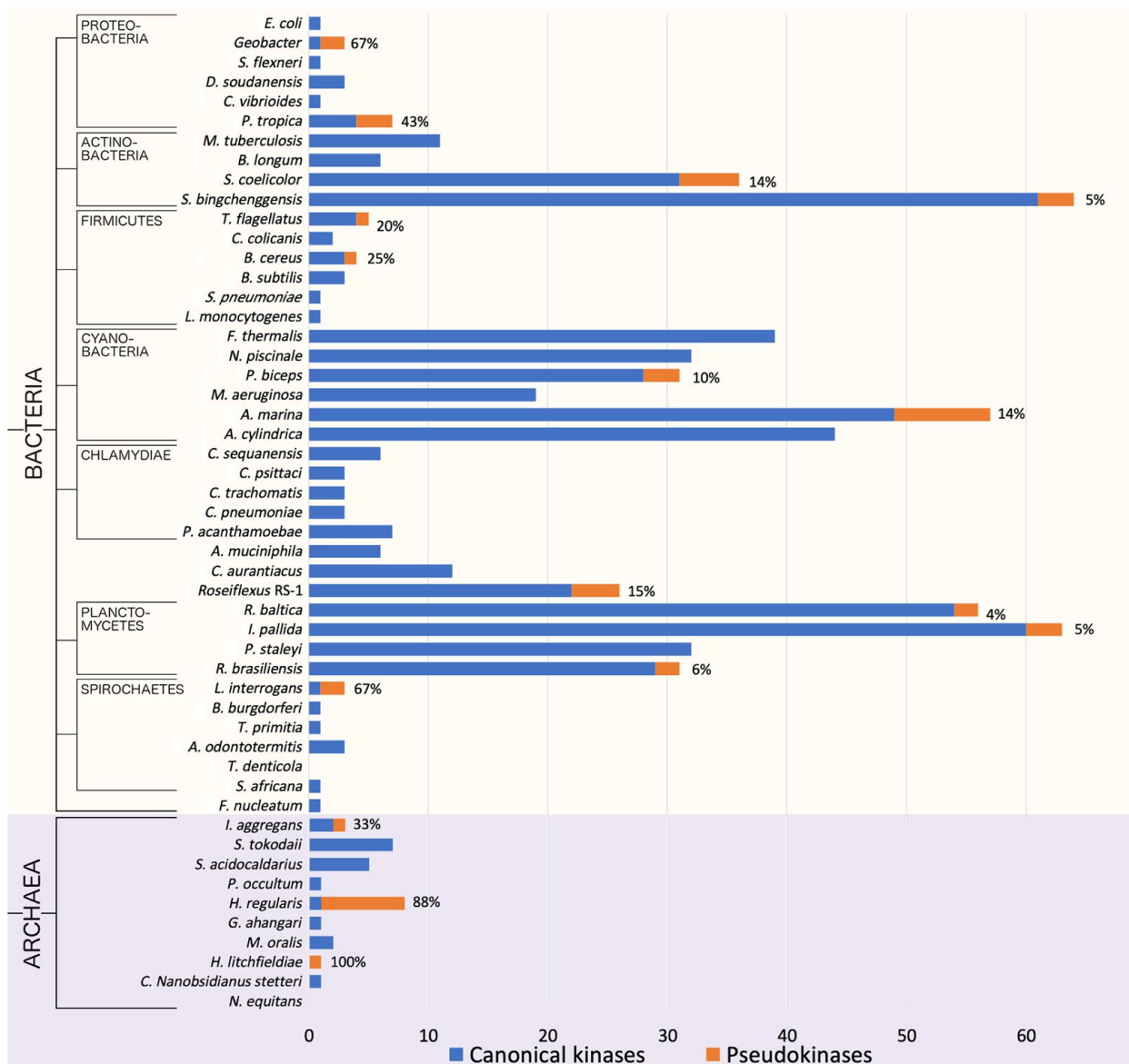


**Fig. 1. Kinome and pseudokinome sizes evaluated in 46 eukaryotic species.** The counts of protein kinases and pseudokinases detectable in various eukaryotic proteomes are shown. Blue bars represent the number of canonical kinases detected in the proteome of each eukaryotic species, and orange bars represent the number of predicted pseudokinases. Percentages indicate the fraction of total kinases from each proteome that were determined to be pseudokinases based on the lack of at least one residue in the catalytic triad. The tree on the left indicates major evolutionary kingdoms and phyla. Reference proteomes were obtained from the UniProt database (59), and protein kinase sequences were extracted and aligned from each reference proteome using ePK sequence profiles. Alignments of kinomes and pseudokinomes from each proteome shown here are provided in data file S1. Counts of kinases and pseudokinases from other eukaryotic proteomes are provided in table S1.

AQ26  $\beta$ -lysine by an arginine within the canonical “VAIK” motif. In addition, we often identify large expansions of pseudokinase families in plant species that are accompanied by the presence of few canonical sequence members. For example, in black cottonwood (*Populus trichocarpa*), we identify one canonical PWNK member containing the  $\beta$ -lysine (“VAWK” motif), along with 15 pseudokinase members containing a “VAWN” motif characteristic of PWNK pseudokinases. The clustering of canonical sequences within pseudokinase families is summarized in table S3, and alignments of these canonical sequences are provided in data file S3.

T3 In addition, our in-depth analysis led to the identification of many previously undefined pseudokinase families (Table 3). These include the plant lysin motif (LysM)-like pseudokinase family and the bacterial HGA motif-containing PknB pseudokinase family, as well as unclassified pseudokinase families that do not readily fall within the identified pseudokinase clusters, which we term

the tyrosine kinase (TK)-, TKL-, STE-, CK1-, AGC-, CAMK-, CMGC-, PknB-, and Other-unclassified pseudokinase families (Fig. 3). As noted previously, pseudokinases in the TKL group appear to have expanded substantially, particularly in nonmetazoan species. Consistently, five distinct fungal-specific pseudokinase families have emerged in the TKL group, of which three (termed Rig1 to Rig3) are currently found predominantly in the species *R. irregularis*, which has a considerably expanded kinome and pseudokinome compared to other fungi (Fig. 1 and fig. S1). In addition, IRAK pseudokinases (part of the TKL group) are massively expanded in plants, which classify into nine unique families. This mirrors the well-known plant-specific expansions of the canonical IRAK (also known as RLK/Pelle) kinase family, which have previously been classified into 65 subfamilies (52). An unexpected finding from our analysis was the detection of over 1200 putative bacterial PknB pseudokinases, which we have classified into 12 distinct clusters. Some bacterial

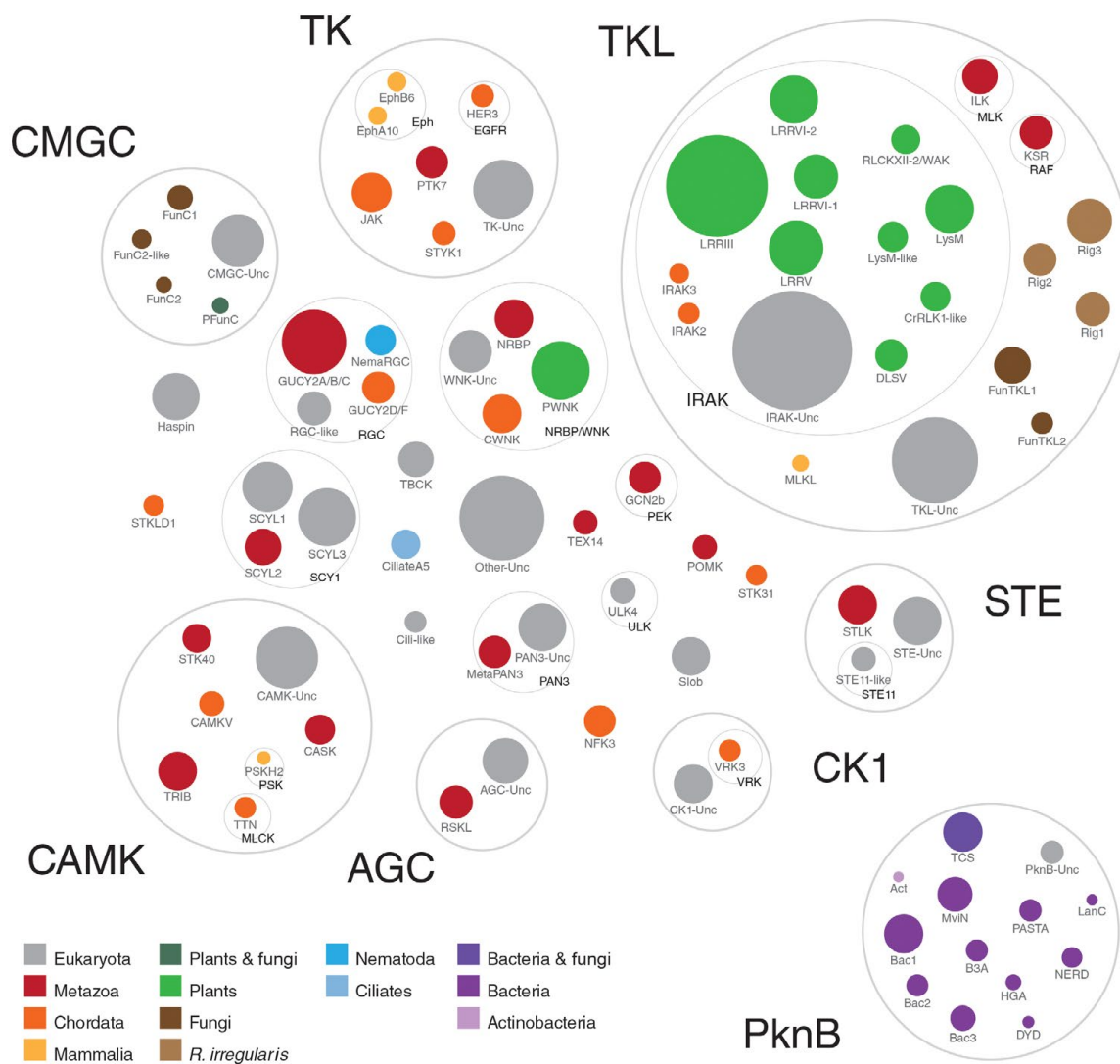


**Fig. 2. Kinome and pseudokinome sizes evaluated in 51 bacterial and archaeal species.** The counts of protein kinases and pseudokinases detectable in various bacterial and archaeal proteomes are shown. Blue bars represent the number of canonical kinases detected in the proteome of each bacterial and archaeal species, and orange bars represent the number of predicted pseudokinases. Percentages indicate the fraction of total kinases from each proteome that were determined to be pseudokinases based on the lack of at least one residue in the catalytic triad. The tree on the left indicates major evolutionary kingdoms and phyla. Reference proteomes were obtained from the UniProt database (59), and protein kinase sequences were extracted and aligned from each reference proteome using diverse sequence profiles of ePKs. Alignments of kinomes and pseudokinomes from each proteome shown here are provided in data file S1. Counts of kinases and pseudokinases from other bacterial and archaeal proteomes are provided in table S1.

pseudokinase clusters are specific to selected bacterial phyla, such as the Actinobacteria-specific PknB pseudokinase family (which we term Act), whereas other families such as the NERD domain-containing PknB pseudokinase family are found more broadly across diverse bacterial phyla. Sequences and alignments for each pseudokinase family identified here are available through the Protein Kinase Ontology (ProKinO, <http://vulcan.cs.uga.edu/prokino/about/browser>) and in the Supplementary Materials (data file S2). In the following sections, we build on the first comprehensive pseudokinase catalog by examining the putative functions of plant IRAK pseudokinases and pseudokinases in *R. irregularis* and in sequenced bacteria.

### A massive, plant-specific IRAK pseudokinase expansion

The plant IRAK kinases have been previously classified into 65 subfamilies (52). We now identify nine unique IRAK pseudokinase families (Fig. 3), which are conserved across multiple plant species (table S2), including two pseudokinase families that resemble the lysin-motif receptor-like (LysM RLK) kinases, termed LysM and LysM-like, and the leucine-rich repeat receptor-like (LRR RLK) pseudokinase families, termed LRRIII, LRRV, LRRVI-1, and LRRVI-2. In addition, we define three “mixed” pseudokinase families that contain domains homologous to multiple previously defined IRAK subfamilies, such as the receptor-like cytoplasmic kinase XII-2 (RLCKXII-2)/WAK family and the *Catharanthus roseus* RLK 1 (CrRLK1)-like family, which are summarized



Novel pseudokinase families:

PFunC	Plant- & fungi-specific CMGC	Bac1; Bac2; Bac3	Bacteria-specific PknB	WNK-Unc; STE-Unc; AGC-Unc
RGC-like		Act	Actinobacteria-specific PknB	IRAK-Unc; TK-Unc; CAMK-Unc;
STE11-like		LanC	LanC domain-containing PknB	CMGC-Unc; PknB-Unc; TKL-Unc;
Rig1; Rig2; Rig3	<i>R. irregularis</i> -specific	MviN	Actinobacteria-specific MviN-like	Other-Unc; CK1-Unc
OoFunTKL	Oomycete- & fungal-specific	HGA	HGA motif-containing PknB	Unclassified
FunTKL	Fungal-specific TKL	DYD	DYD motif-containing PknB	
FunC1; FunC2	Fungi-specific CMGC	B3A	Beta3-alanine-containing	
FunC2-like	Fungi-specific CMGC-2-like	NERD	NERD domain-containing	
Cili-like	Ciliate kinase-like	PASTA	PASTA domain-containing	
NemaRGC	Nematode-specific RGC	TCS	Two-component system PknB	
PWNK	Plant-specific WNK			
CWNK	Chordate-specific WNK			
MetaPAN3	Metazoan-specific PAN3			

AQ62 **Fig. 3. A new classification of pseudokinase families.** The organization of pseudokinase sequences into distinct, evolutionarily related pseudokinase families is depicted. Pseudokinase sequences were classified into families based on unique patterns of amino acid conservation and variation. Each colored circle represents a distinct pseudokinase family, which is colored according to the taxonomic group(s) in which it is found. The size of each circle represents the relative size of the corresponding pseudokinase family in log scale, and the distances between circles reflect the approximate sequence divergence between families as determined by hidden Markov models (HMM)-to-HMM distances. Kinase groups and families from the human kinome classification (18) are depicted by gray circles and labeled in bold font. Sequences and alignments for each pseudokinase family shown here are provided in data file S2.

in table S2. The widespread distribution of these new plant pseudokinase families across diverse plant species is summarized in table S4.

To specifically examine the evolutionary expansion of IRAK pseudokinases in plants, we constructed a rooted phylogenetic tree

of both canonical and pseudokinase IRAK sequences using diverse non-IRAK TKL sequences and non-TKL sequences as an outgroup. The plant IRAK pseudokinase families form distinct clades in the phylogenetic tree that distinguish them from metazoan IRAK pseudokinase

AQ67

AQ68

**Table 3. List of newly identified pseudokinase families.** Representative members of previously unidentified pseudokinase families are listed. WAK, wall-associated kinase; CAMK, calcium/calmodulin-dependent protein kinase; RGC, receptor guanylyl cyclase.

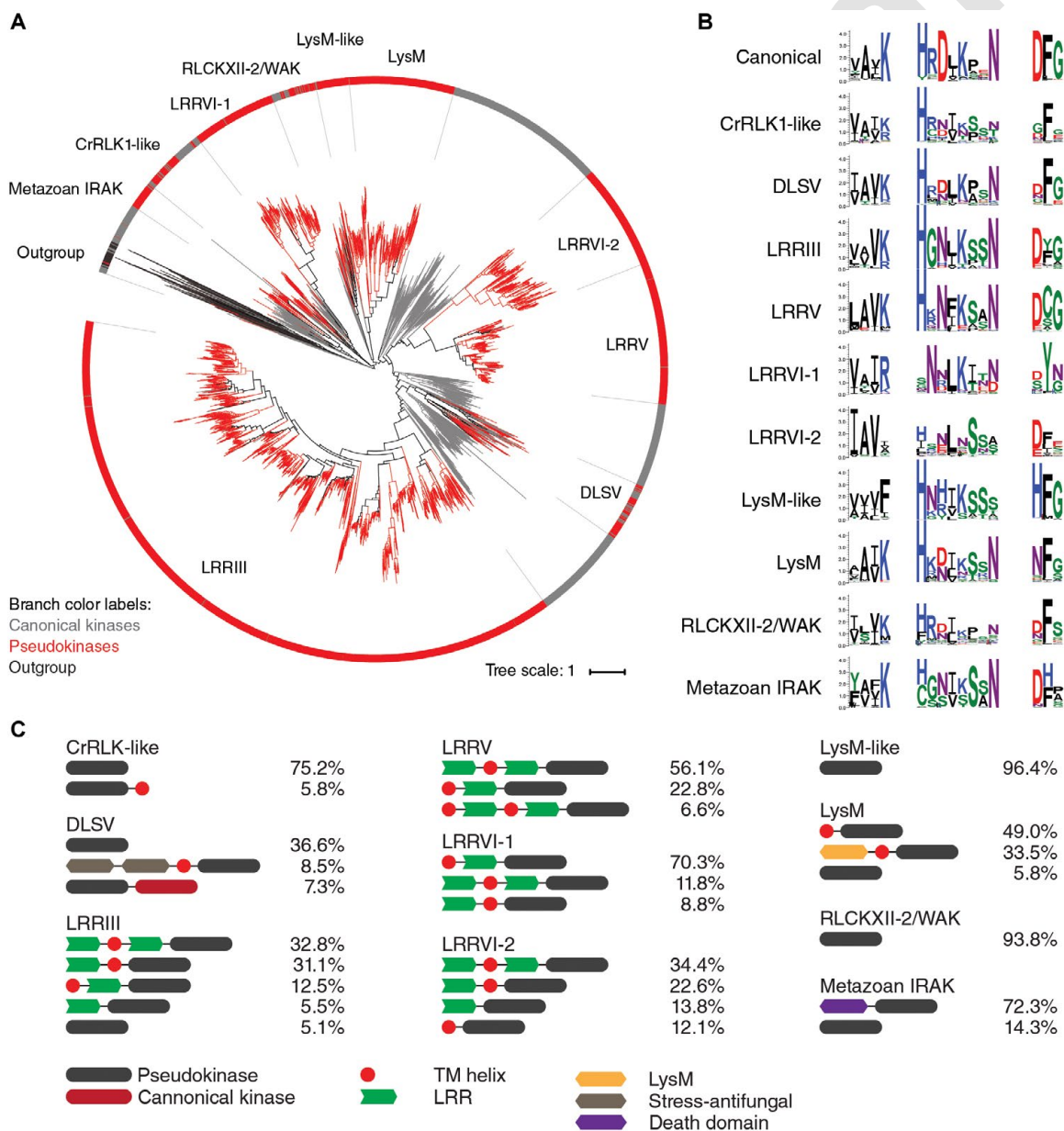
PsK family	Kinase family	Kinase group	Example member(s)
TK-Unclassified		TK	<i>Camponotus floridanus</i> putative tyrosine-protein kinase Wsck (E2ACX7_CAMFO)
LysM	IRAK	TKL	<i>A. thaliana</i> LYK5/LysM-containing receptor-like kinase 5/At2g33580 (LYK5_ARATH)
LysM-like	IRAK	TKL	<i>A. thaliana</i> protein kinase superfamily protein/At3g57120 (Q8VYG5_ARATH)
DLSV	IRAK	TKL	<i>A. thaliana</i> cysteine-rich receptor-like protein kinase 45/Cysteine-rich RLK45/At4g11890 (CRK45_ARATH)
RLCKXII-2/WAK	IRAK	TKL	<i>A. thaliana</i> nonfunctional pseudokinase ZED1/hopZ-ETI-deficient 1/At3g57750 (ZED1_ARATH)
CrRLK1-like	IRAK	TKL	<i>Cajanus cajan</i> receptor-like protein kinase ANXUR2/KK1_043523 (A0A151QYL0_CAJCA)
LRRIII	IRAK	TKL	<i>A. thaliana</i> PRK1/Pollen receptor-like kinase 1/At5g35390 (PRK1_ARATH)
LRRV	IRAK	TKL	<i>A. thaliana</i> SUB/STRUBBELIG/At1g11130 (SUB_ARATH)
LRRVI-1	IRAK	TKL	<i>A. thaliana</i> leucine-rich repeat protein kinase family protein/MUA22.21/At5g14210 (A0A178US29_ARATH)
LRRVI-2	IRAK	TKL	<i>A. thaliana</i> MDIS1/Male discoverer 1/At5g45840 (MDIS1_ARATH)
Rig1		TKL	<i>R. irregularis</i> Rad53p/RirG_173180 (A0A015K1H5_9GLOM)
Rig2		TKL	<i>R. irregularis</i> Skt5p/RirG_180750 (A0A015ISZ5_9GLOM)
Rig3		TKL	<i>R. irregularis</i> Ssk22p/RirG_232660 (A0A015IBX1_9GLOM)
FunTKL1		TKL	<i>Rhizoctonia solani</i> uncharacterized protein/RSAG8_12895 (A0A066UWY2_9HOMO)
FunTKL2		TKL	<i>Penicillium subrubescens</i> uncharacterized protein/PENSUB_13048 (A0A1Q5SUF1_9EURO)
IRAK-Unclassified		TKL	<i>A. thaliana</i> BIR2/BAK1-interacting receptor-like kinase 2/At3g28450 (BIR2_ARATH)
STE11-like		STE	<i>A. thaliana</i> protein kinase superfamily protein/At2g40580 (O22877_ARATH)
STE-Unclassified		AGC	<i>Homo sapiens</i> uncharacterized protein (A0A1B0GUL7_HUMAN); <i>Pan troglodytes</i> uncharacterized protein (A0A2I3RK43_PANTR)
AGC-Unclassified		AGC	<i>P. tetraurelia</i> hypothetical protein/GSPATT00021006001 (A0DWE7_PARTE)
CAMK-Unclassified		CAMK	<i>P. tetraurelia</i> uncharacterized protein/GSPATT0003535001 (A0E5V5_PARTE)
PFunC		CMGC	<i>Sorghum bicolor</i> uncharacterized protein/SORBL_3005G165300 (A0A1Z5RK15_SORBI)
FunC1		CMGC	<i>Penicillium patulum</i> uncharacterized protein/ PGRI_024730 (A0A135LI09_PENPA)
FunC2		CMGC	<i>Penicillium subrubescens</i> SRSF protein kinase 3/PENSUB_10992 (A0A1Q5T6W4_9EURO)
FunC2-like		CMGC	<i>Coprinopsis cinerea</i> CMGC/SRPK protein kinase/CC1G_09673 (A8P9H1_COPC7)
NemaRGC	RGC	Other	<i>Caenorhabditis elegans</i> receptor-type guanylate cyclase gcy-1/AH6.1 (GCV1_CAEEL)
RGC-like	RGC	Other	<i>Amphimedon queenslandica</i> guanylate cyclase (A0A1X7VH93_AMPQE)
CWNK	WNK	Other	<i>Homo sapiens</i> WNK1/protein kinase with no lysine 1 (WNK1_HUMAN)
PWNK	WNK	Other	<i>A. thaliana</i> AtWNK1/protein kinase with no lysine 1/At3g04910 (WNK1_ARATH)
WNK-Unclassified	WNK	Other	<i>Caenorhabditis elegans</i> serine/threonine-protein kinase WNK/Protein kinase with no lysine 1/C46C2.1 (WNK_CAEEL)
Cili-like		Other	<i>Neurospora crassa</i> mitotic spindle checkpoint component mad3/NCU10043 (V5IRN2_NEUCR)
MetaPAN3		PknB	<i>Homo sapiens</i> PAN3/Poly(A)-nuclease deadenylation complex subunit 3 (PAN3_HUMAN)
Bac1		PknB	<i>Streptomyces aurantiacus</i> JA 4570 uncharacterized protein/STRAU_1734 (S3ZNQ3_9ACTN)
Bac2		PknB	<i>Frankia</i> sp. E15c serine/threonine protein kinase/UG55_100447 (A0A166T1F0_9ACTN)
Bac3		PknB	<i>Trichodesmium erythraeum</i> (strain IMST01) serine/threonine protein kinase with TPR repeats/Tery_4781 (Q10VJ1_TRIEI)
Act		PknB	<i>Streptomyces</i> sp. NTK 937 aminoglycoside phosphotransferase/DT87_12690 (A0A069K2D6_9ACTN)
MviN		PknB	<i>Mycobacterium shimoidei</i> transmembrane serine/threonine-protein kinase D PknD/STPK D/ BHQ16_10175 (A0A1E3TG29_MYCSH)
LanC		PknB	<i>Streptomyces vietnamensis</i> uncharacterized protein /SVTN_34785 (A0A0B5I836_9ACTN)
DYD		PknB	<i>Frankia</i> sp. (strain EAN1pec) nucleic acid-binding OB-fold tRNA/helicase-type/Franean1_4264 (A8L9E5_FRASN)
HGA		PknB	<i>Actinosynnema mirum</i> (strain ATCC 29888) uncharacterized protein/Amir_3792 (C6WDS3_ACTMD)
B3A		PknB	<i>Lentisphaera araneosa</i> HTCC2155 serine/threonine-protein kinase/LNTAR_17493 (A6DFI7_9BACT)
PASTA		PknB	<i>Firmicutes bacterium</i> CAG:94 PASTA domain protein/BN815_00934 (R6ZDX7_9FIRM)
NERD		PknB	<i>Singulisphaera acidiphila</i> (strain ATCC BAA-1392) nuclease-like protein/protein kinase family protein/Sinac_1959 (L0DAB9_SINAD)
TCS		PknB	<i>Rhodospirella islandica</i> two-component signal transduction/RISK_006087 (A0A0J1B5P6_RHOIS)



F4 sequences (Fig. 4A). In addition, metazoan IRAK members are cytoplasmic and typically contain a death domain N-terminal to the kinase domain (Fig. 4C), which is not observed in any plant IRAK kinases. We note that the three mixed pseudokinase families we identified (RLCKXII-2/WAK, CrRLK1-like, and DLSV) form distinct monophyletic clades in the phylogenetic tree, indicating close homology and common descent, despite their homology to multiple IRAK subfamilies. In addition, the IRAK pseudokinases are generally divergent from canonical plant IRAK members, as shown by the long branch lengths separating the pseudokinase family clades (Fig. 4A, red lines) from the canonical IRAK sequences (Fig. 4A,

gray lines). In some pseudokinase families, several canonical sequences cluster within pseudokinase family clades (namely, RLCKXII-2/WAK, CrRLK1-like, DLSV, and LRRIII and LRRV families), indicating that these pseudokinase families also have very close, likely catalytically active, homologs.

We next examined the relative amount of degradation of the canonical catalytic motifs and the domain organizations of the major IRAK pseudokinase families, allowing us to expand on the potential functions of some IRAK pseudokinases. The LysM pseudokinase family shares sequence homology in the kinase domain with known catalytically active LysM RLKs and conserves a similar domain



**Fig. 4. Plant-specific IRAK pseudokinase families.** (A) Phylogenetic tree of catalytically active and pseudokinase members of the IRAK family. The nine plant IRAK pseudokinase families are labeled, and IRAK pseudokinase sequences are shown in red. Canonical IRAK sequences are shown in gray. Outgroup sequences are shown in black. (B) Sequence logos of catalytic motifs for IRAK pseudokinase families. (C) Unique domain structures observed in plant IRAK pseudokinase families. The most common domain structures observed in each family are shown (occurring >5%), with frequencies of each domain structure indicated.



architecture, which is characterized by an intracellular kinase domain, and one or more extracellular LysM domains (Fig. 4C). In plants, LysM RLKs play diverse sensing functions, recognizing chitin oligosaccharides in plant defense responses toward fungi (67), as well as binding peptidoglycans on bacterial cell walls to aid recognition of symbiotic bacteria (68). Recent studies of two LysM pseudokinases, MtNFP and LjNFR5 from *Medicago*, demonstrate the importance of MtNFP and LjNFR5 during the *Rhizobium* preinfection response and to the specificity of *Rhizobium*-legume symbiosis, respectively. Despite their lack of catalytic activity, these LysM pseudokinases are believed to contribute to their appropriate signaling pathways via interactions with other active LysM RLKs (69). We also detect another distinct family, which we term the LysM-like pseudokinase family, which shares sequence homology in the kinase domain with active LysM RLKs but lacks LysM and transmembrane (TM) domains, suggesting a uniquely cytoplasmic function for this family (Fig. 4C). Moreover, the LysM and LysM-like families share different patterns of residue conservation, as identified in our pattern-based classification, and they form distinct clades in the phylogenetic tree (Fig. 4A), suggesting that they likely have divergent functions.

The LRRV pseudokinase family comprises the Strubbelig (SUB) receptor family, which consists predominantly of pseudokinases, with nine members in *Arabidopsis thaliana* alone (70). The best characterized member of this family, SUB, plays roles in organ development and cellular morphogenesis; however, the mechanism by which it contributes to cell signaling despite a lack of catalytic activity and the functions of other LRRV members are not yet known (70, 71). Although the domain organizations for most IRAK pseudokinase families such as LRRV are rather well conserved, pseudokinases in the DLSV family have diverse domain architectures, indicating that a common pseudokinase domain has coevolved with a variety of different protein domains to diversify biological function. For example, DLSV pseudokinases homologous to the DUF26 IRAK subfamily contain extracellular domains associated with salt-stress and antifungal responses (Fig. 4C) (72, 73). Although DUF26 kinases have been associated with plant-specific functions in reactive oxygen species/redox signaling and stress adaptation (74), to our knowledge, pseudokinase complements of DUF26 kinases have not been described previously. Other DLSV pseudokinases exhibit tandem pseudokinase and canonical kinase domains (Fig. 4C). Previously described IRAK pseudokinases in plants include MDIS2 (MRH1) (75), ZED1 (51), RSK1 (76), BSK8 (77), BIR2 (49), SOBIR1 (78), and CRN (79), and their placement in the expanded IRAK pseudokinase classification is described in table S5.

### Expansion of TKL pseudokinases in *R. irregularis*

We identified and analyzed three previously unidentified fungal pseudokinase families (termed Rig1, Rig2, and Rig3), which are currently composed predominantly of sequences from a single species, the commercially important soil inoculant *R. irregularis* (Fig. 3). This organism has a highly expanded proteome when compared to other fungal species, including an expanded kinome (65, 80). The *R. irregularis* fungal kinome comprises ~2.6% of the entire proteome, a larger proportion than is observed for any other fungal kinome analyzed (fig. S1). Thirty-two percent of these kinases have pseudokinase domains (Fig. 1) that most closely resemble the TKL kinases. Several lines of evidence suggest that the genes coding these pseudokinases are expressed at the protein level (65, 81), suggesting

that protein pseudokinases must contribute to an important biological function in *R. irregularis*.

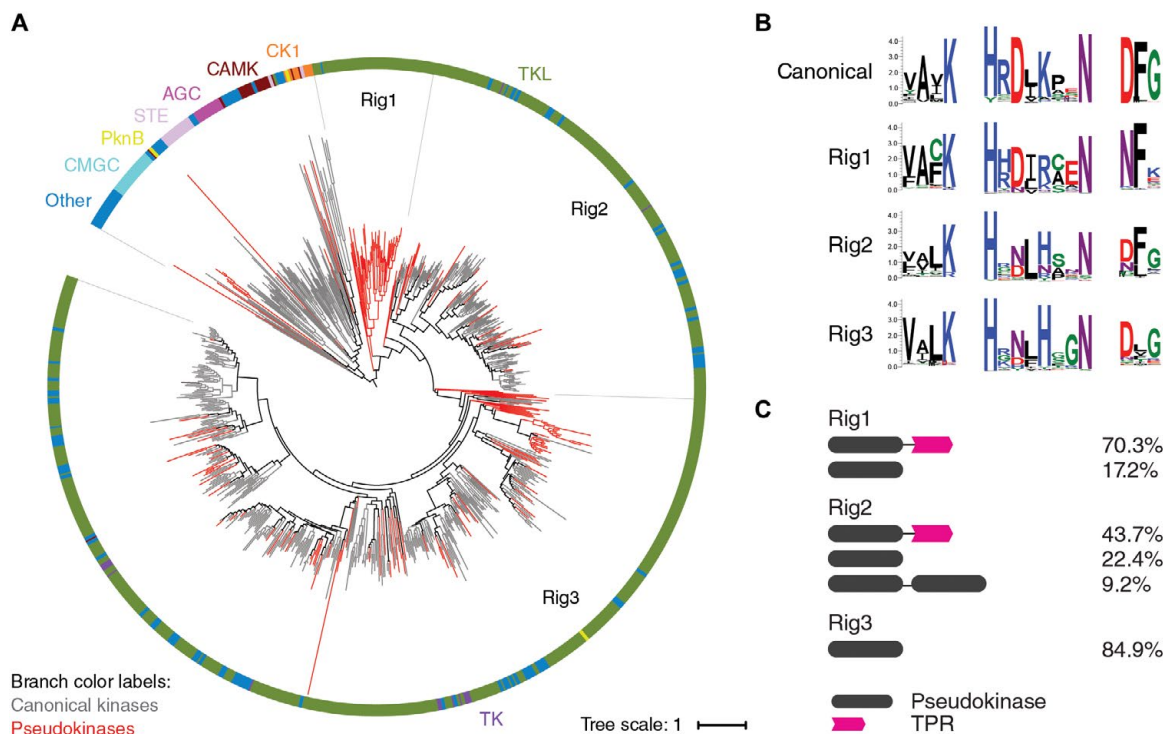
Sequence comparisons to known TKL kinase families demonstrated that *R. irregularis*-specific TKL kinases are divergent from canonical TKL families and are most homologous to the leucine rich repeat kinase (LRRK) family of TKLs. LRRKs are intracellular kinases distinct from IRAK LRR RLKs and are conserved in metazoans and expanded in the slime-mold *Dictyostelium discoideum* but are otherwise absent in fungi. To understand the evolutionary events that led to the expansion of these pseudokinase families, we conducted a phylogenetic analysis of the entire *R. irregularis* kinome, which, like the human kinome (18), consists of seven major ePK groups. *R. irregularis* kinase sequences from the AGC, CMGC, STE, CAMK, and CK1 groups form mostly separate monophyletic clades in a phylogenetic tree of the *R. irregularis* kinome, whereas TK sequences are clustered within various groupings of TKLs (Fig. 5A). The *R. irregularis* kinome additionally includes PknB-like kinase sequences, which are typically associated with bacteria. However, as reflected by the phylogenetic tree (Fig. 5A), TKL sequences compose most of the *R. irregularis* kinome. Thus, the expanded kinome of *R. irregularis* can be attributed to a substantial expansion among TKL kinases. Furthermore, we found that 166 of the 183 pseudokinases (90%) identified in *R. irregularis* are TKL-like, indicating that *R. irregularis* pseudokinases emerged primarily from within the TKL group. Of the three distinct pseudokinase families, Rig1 forms a monophyletic group composed entirely of pseudokinases, whereas Rig2 and Rig3 cluster into clades including both pseudokinase and canonical kinase members. Rig2 and Rig3 are the most diverse pseudokinase families in *R. irregularis*, and both cluster with canonical sequences from the TK and “Other” groups, based on the human kinome classification (18).

Further analysis of the domain organization in these pseudokinase sequences revealed that Rig1 and Rig2 pseudokinases often have an additional putative tetratricopeptide (TPR) domain C-terminal to the pseudokinase domain, whereas Rig3 pseudokinases are single-domain pseudokinases (Fig. 5C). TPR repeats are short structural motifs that are classically involved in mediating protein-protein interactions crucial for cell signaling. Apart from the TPR repeat regions, no additional domains were found in the pseudokinase members of the three families.

### Conservation of PknB-like pseudokinases in bacteria

We identified 12 unique families of bacterial pseudokinases that are most closely related to the PknB group of canonical prokaryotic protein kinases (Fig. 3) and have been classified and named on the basis of their conserved domains, taxonomic specificity, and/or their similarity to previously identified kinases. Three bacterial pseudokinase families (DYD, HGA, and B3A) were named on the basis of the unique conservation of noncanonical catalytic motifs. To examine the evolutionary relationships between these families, we built a phylogenetic tree of the 1145 PknB pseudokinase sequences identified in this study combined with a representative set of canonical PknB kinases (Fig. 6A). Each bacterial pseudokinase family falls into a distinct clade and mostly segregates away from canonical kinase sequences.

Analysis of the catalytic motifs for each PknB pseudokinase family shows that different PknB pseudokinase families diverge in the canonical catalytic motifs in a variety of ways (Fig. 7B). For example, the MviN and PASTA families exhibit the most extreme degeneration



**Fig. 5. *R. irregularis*-specific TKL pseudokinase families.** (A) Phylogenetic tree of the *R. irregularis* kinome. Canonical kinase branches are colored in gray, and pseudokinases are in red. Major kinase groups are labeled using different colors in the outer circle. The three major *R. irregularis*-specific pseudokinase families are labeled as Rig1, Rig2, and Rig3. (B) Sequence logos of catalytic motifs for Rig1, Rig2, and Rig3 pseudokinase families. (C) The most common domain structures observed in Rig pseudokinase families are shown (occurring >5%), with frequencies of each domain structure indicated.

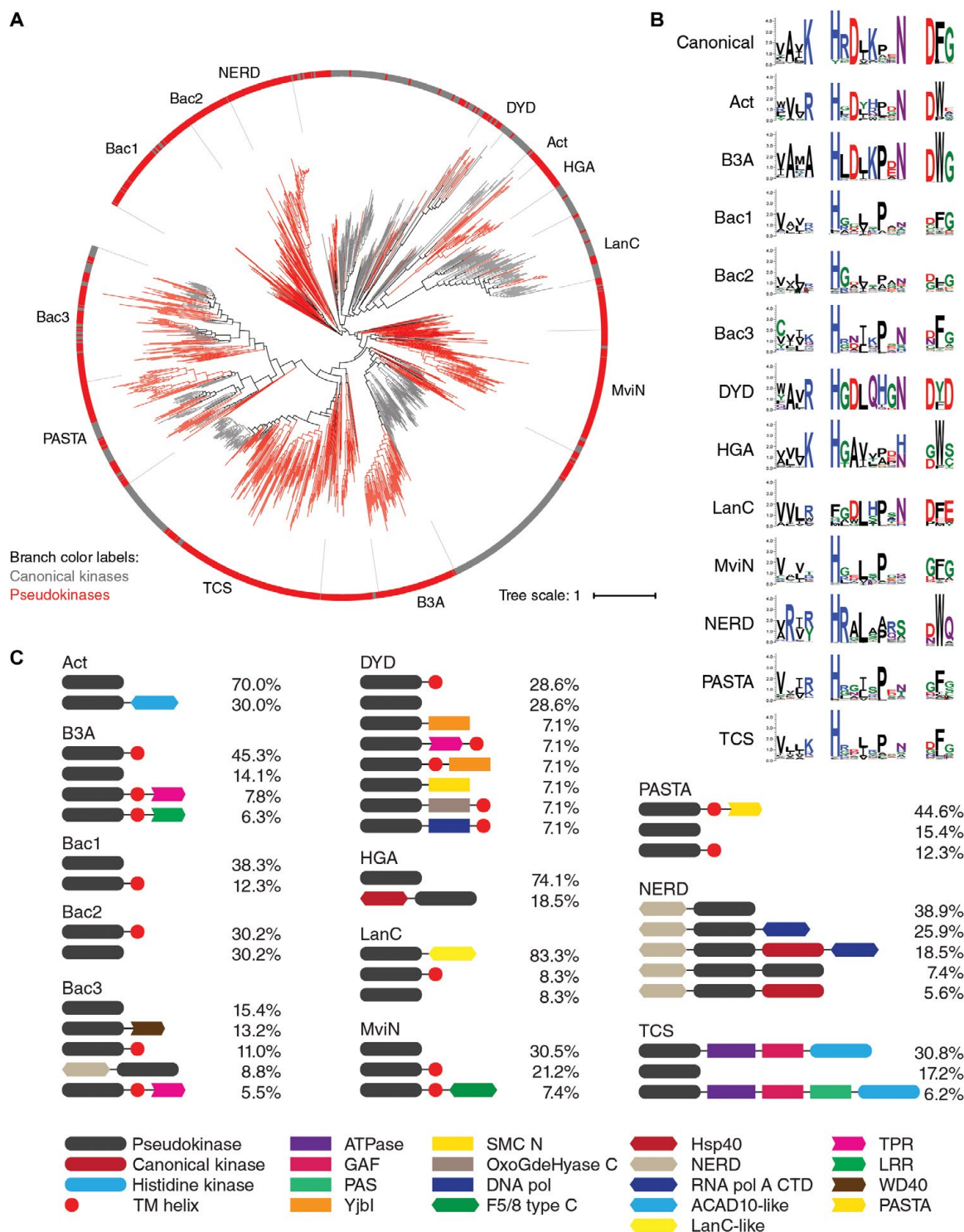
of catalytic residues and lack all three canonical residues associated with catalysis ( $\beta$ -lysine, HRD-aspartate, and DFG-aspartate), in addition to the conserved magnesium-binding asparagine located at the end of the catalytic loop HRDXXXXN motif. Conversely, many bacterial pseudokinase families appear to conserve the catalytic aspartates of the HRD and DFG motifs, as well as the magnesium-binding asparagine residue; this is the case for the Act, DYD, B3A, and LanC pseudokinase families. Of these, the Act, DYD, and LanC families all have a chemically comparable arginine residue in place of the ATP-binding  $\beta$ -lysine and thus might still retain ATP binding and catalytic functions, as demonstrated for canonical kinases such as Aurora A (82).

Analysis of the common domain organizations of bacterial PknB pseudokinase families reveals that, like eukaryotic pseudokinases, bacterial pseudokinases also co-occur with protein signaling domains involved in a wide range of biological functions (Fig. 6C). Two families in particular, termed NERD and TCS, have highly conserved domain architectures of particular interest. For example, the domain architecture of NERD is characterized by an N-terminal NERD domain, which is predicted to function as a nuclease (83), followed by a PknB pseudokinase domain and often a C-terminal canonical (catalytically active) PknB kinase domain. This multidomain architecture is also observed in PglW, a constituent of the *S. coelicolor* A(3)2 phage growth limitation system (84). The active kinase domain of the PglW is catalytically competent (85), whereas the function of the pseudokinase domain remains unknown. In addition, we identified putative protein domains that resemble the C-terminal domain of the bacterial RNA polymerase  $\alpha$ -subunit in ~45% of the NERD family members; this domain classically functions in DNA bind-

ing and protein-protein interactions in bacterial RNA polymerases (86). The co-occurrence of these nucleic acid-associated domains with pseudokinase domains suggests that the NERD pseudokinase family may be involved in signaling pathways relevant to transcriptional regulation. We also note the unique domain structure within the TCS family, which has clear orthologs in both bacteria and fungi (Fig. 3 and table S2) and contains a pseudokinase domain that often co-occurs with other protein domains typically found in two-component signaling systems involving histidine and aspartate phosphorylation. Nearly 40% of TCS family pseudokinases contain an N-terminal pseudokinase domain followed by an adenosine triphosphatase domain, a GAF domain, and a C-terminal histidine kinase domain (Fig. 6C). This domain architecture is reminiscent of two-component system proteins that have been previously identified in bacterial species from the Cyanobacteria, Proteobacteria, and Spirochaete phyla, as well as fungal species such as *Candida albicans* and *Schizosaccharomyces pombe* (87–92). These proteins have been associated with a wide range of functions, including nitrogen metabolism and glycolipid synthesis in *Anabaena* sp. PCC7120, hyphal development and pathogenicity in *C. albicans*, and cell cycle regulation and oxidative stress response in *S. pombe* (87, 91, 93–101). To our knowledge, our analysis is the first to reveal pseudokinase domains that are likely to be associated with two-component signaling in bacteria.

### Sequence and structural basis for pseudokinase evolutionary divergence: A case study on two plant IRAK pseudokinase families

Conformational flexibility in pseudokinases has permitted the structurally conserved ancestral protein kinase fold to be “repurposed”

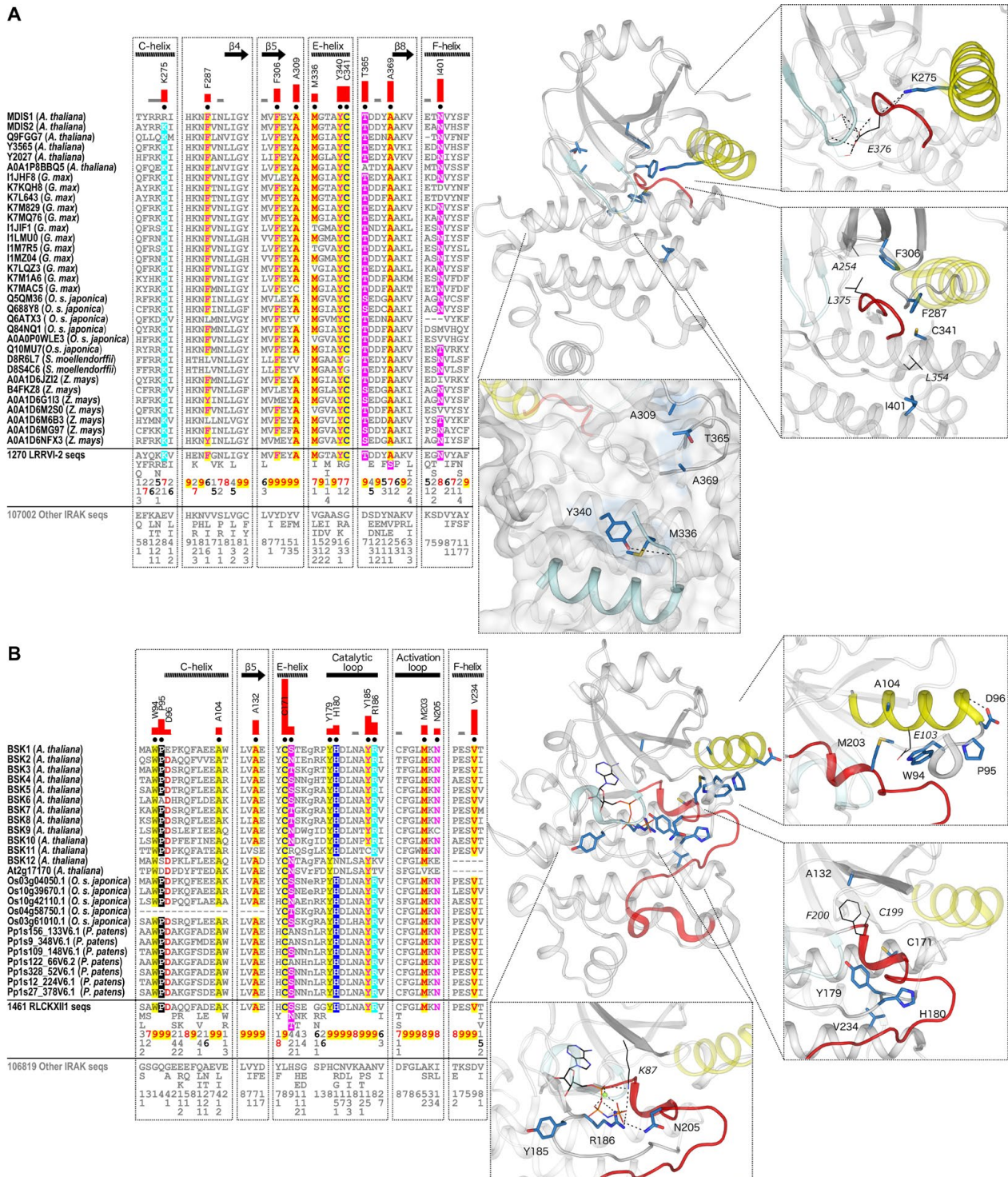


**Fig. 6. Bacterial PknB pseudokinase families.** (A) Phylogenetic tree of PknB canonical kinases and pseudokinases. The 12 PknB-related pseudokinase families are labeled and shown in red branches. Representative canonical PknB kinases are shown in gray. (B) Sequence logos of catalytic motifs in PknB pseudokinase families. (C) Unique domain structures observed in bacterial pseudokinase families. The most common domain structures observed in each family are shown (occurring >5%), with percentages indicating the frequency of each domain structure.

for multiple cellular signaling roles, including the evolution of new ways through which to bind and modulate cellular targets. Using the LRRVI-2 and RLCKXII-1 pseudokinase families as examples,

we evaluated how evolution has constrained sequences in different pseudokinase families to disrupt canonical ATP binding and constrain kinase domain conformations in a multitude of ways that





**Fig. 7. IRAK pseudokinase-specific features contribute to unique conformations in key catalytic regions.** (A) LRRVI-2 pseudokinase family-specific sequence motifs. In the alignment, columns are highlighted where amino acids are highly conserved in LRRVI-2 pseudokinase family sequences and nonconserved and/or biochemically dissimilar in other IRAK sequences. Red bar lengths quantify the degree of divergence between LRRVI-2 and other IRAK sequences. Column-wise amino acid and insertion/deletion frequencies are indicated in integer tenths, where a "5" indicates an occurrence of 50 to 60% in the given (weighted) sequence set. Columns used by the Bayesian partitioning procedure to sort LRRVI-2 sequences from other IRAK sequences are marked with black dots. Kinase secondary structures are annotated above the alignment.

(B) RLCKXII-1 pseudokinase family-specific sequence motifs. A structure of BSK8 from *A. thaliana* (PDB 4I94) was used to analyze RLCKXII-1-specific sequence motifs.



serve to abolish catalytic activity. This leads us to propose previously unknown molecular functions that may have evolved in LRRVI-2 pseudokinases through the selection of unique motifs on the surface of the protein.

The LRRVI-2 pseudokinase family includes the previously described pseudokinase, MDIS2, which interacts with the plant potassium channel AKT2 (75) associated with root hair formation (102). However, little is known about the molecular functions of other pseudokinases in this family. Using the crystal structure of a LRRVI-2 pseudokinase from *Zea mays* (GRMZM2G135359), we examined the structural role of LRRVI-2-specific motifs. ATP binding in the GRMZM2G135359 structure appears to be completely inhibited due to the complete obstruction of the ATP-binding site by the activation loop. This activation loop conformation is stabilized by a LRRVI-2-specific lysine (Lys<sup>275</sup>) that replaces the ATP-binding C-helix glutamate (Fig. 7A) and hydrogen bonds to the backbone of the activation loop. A glutamate in the activation loop “DLE” motif, which replaces the canonical DFG motif, hydrogen bonds to the glycine-rich loop to occlude ATP binding. The inhibitory activation loop conformation is additionally stabilized by hydrophobic interactions between LRRVI-2-specific residues including a phenylalanine in the gatekeeper position on the  $\beta 5$  strand (Phe<sup>306</sup>), a phenylalanine in the  $\alpha C$ - $\beta 4$  loop (Phe<sup>287</sup>), and a cysteine in the E-helix (Cys<sup>341</sup>). These residues form hydrophobic interactions with Ala<sup>254</sup> (which replaces the ATP-binding  $\beta 3$ -lysine) and with Leu<sup>375</sup> (which replaces the DFG phenylalanine). Together, these hydrophobic interactions appear to stabilize the C-helix in an inactive, outward conformation and the activation loop in an autoinhibitory conformation that occludes ATP binding. In addition, a LRRVI-2 family-specific asparagine replaces the canonical F-helix aspartate, which typically participates in a switch-like mechanism and stabilizes the active conformation of the kinase domain by forming key hydrogen bonds with the catalytic loop backbone (103).

The F-helix aspartate is typically conserved as an asparagine in LRRVI-2 pseudokinases, although the GRMZM2G135359 structure has a hydrophobic isoleucine at this position, and other LRRVI-2 members are noted to have a valine, methionine, or aspartate (Fig. 7A). Mutation of the F-helix aspartate residue to an asparagine or a leucine in canonical kinases such as Aurora A abrogates activity (103), thus, the substitution of the F-helix aspartate to an asparagine or hydrophobic residue is predicted to inactivate LRRVI-2 pseudokinases. Likewise, variations in the catalytic loop (replacement of the HRD motif by LRN motif) may contribute to the observed inactive structural conformation in LRRVI-2. In addition, we found that several LRRVI-2 family-specific motifs occur well outside of the catalytically important regions, including the surface of the protein near the N- and C-terminal tails. We note one such example in the GRMZM2G135359 structure, where LRRVI-2-specific residues appear to dock the I-helix and C-terminal tail onto the backside of the kinase domain. Specifically, a methionine (Met<sup>336</sup>) and a tyrosine (Tyr<sup>340</sup>) on the E-helix tether the C-tail through hydrophobic and hydrogen bonding interactions, respectively (Fig. 7A). Another cluster of LRRVI-2-specific, surface-exposed residues (A309, T365, and A369) may also participate in tethering the C-tail and extend this interaction to the hinge region of the kinase domain.

To further investigate the evolutionary basis for IRAK pseudokinase functional specialization, we next quantified the evolutionary

constraints imposed on the RLCKXII-1 family of pseudokinases, which comprise the brassinosteroid signaling kinases (BSKs). BSKs are involved in regulating plant growth and physiology in response to brassinosteroid hormone signals. The crystal structure of BSK8 has been determined (77) and shown to bind the nonhydrolyzable ATP analog adenylyl-imidodiphosphate (AMP-PNP), despite the atypical DFG motif (CFG in BSK8) conformation. In addition to an unusual glycine-rich loop structure and the presence of a conserved small amino acid at the gatekeeper position (Ala<sup>132</sup>) (77), our studies also reveal family-specific variations in both the active site (Tyr<sup>185</sup>, Arg<sup>186</sup>, and Asn<sup>205</sup>) and in allosteric regions such as the F-helix aspartate (Val<sup>234</sup>), which provide clues to the unusual mode of AMP-PNP binding in RLCKXII-1 (Fig. 7B). Family-specific replacement of the canonical HRD motif histidine and arginine in the catalytic loop (Tyr<sup>179</sup> and His<sup>180</sup>) facilitates unique hydrogen bonding and hydrophobic interactions that stabilize the activation loop in a unique inactive conformation (Fig. 7B). Other RLCKXII-1 features contribute to unique inactive conformations of the C-helix via hydrophobic packing interactions with the ATP-binding C-helix glutamate (Glu<sup>103</sup>, Ala<sup>104</sup>, Met<sup>203</sup>, and Trp<sup>94</sup>) and by promoting a unique secondary structure of the  $\beta 3$ - $\alpha C$  loop and C-helix (Pro<sup>95</sup> and Asp<sup>96</sup>). These findings add to the seemingly limitless ways in which ePK superfamily members can evolve new sequence features to affect kinase conformations and to ultimately modulate signaling outputs from the kinase domain.

## DISCUSSION

Using a large-scale bioinformatics analysis, we have considerably expanded the classification of pseudokinases. We identified a total of 86 putative pseudokinase families and demonstrated that pseudokinases are present across the tree of life. Our analysis strongly suggests that pseudokinases are polyphyletic, emerging through numerous events during the course of protein kinase evolution, presumably to fulfill different biological niches and signaling roles through noncatalytic functions, the broad extent of which is revealed by our analysis.

Pseudokinases have evolved in all the major ePK groups; however, the TKL group is particularly enriched with pseudokinases, largely due to expansions of TKL kinases in plants and fungi. These TKL group expansions occur in the IRAK family in plants, whereas TKL expansions in fungi, including *R. irregularis*, comprise distinct pseudokinase families unrelated to any known TKL families in other organisms, corroborating previous descriptions of “unclassifiable” kinases in fungi (104, 105). Why have TKLs in particular been selected during evolution for such marked kinome and pseudokinome expansions in both plants and fungi? One possible explanation is the lack of TKs in these organisms, which, in metazoa, evolved and duplicated to play crucial phosphotyrosine-dependent roles in multicellular signal transduction (106). Whereas TKs compose most of the receptor protein kinase repertoire in metazoans, the IRAK family of TKLs comprises a receptor kinase-like repertoire in plants, and thus, the expansion of IRAK kinases and pseudokinases in plants may be analogous to the expansion of receptor TKs in metazoa. In line with this view, LysM pseudokinases in *Medicago* are believed to contribute to *Rhizobium* interactions by interacting with active LysM RLK members, which is reminiscent of metazoan tyrosine pseudokinases, such as HER3, which allosterically modulates closely related, active EGFR family members (36, 69). Thus, the expansion

of IRAK pseudokinases in plants mirrors the expansion of canonical IRAK kinases, perhaps due to regulatory interactions between coevolved kinases and pseudokinases. Mechanistically, plant IRAK kinases are known to play vital roles in plant-fungi interactions, both during symbiotic interactions and pathogen defense, suggesting that the expansion of fungal TKL kinases and pseudokinases may have arisen due to a close symbiotic coevolution. *R. irregularis* participates in arbuscular mycorrhizal symbiotic relationships with more than two-thirds of all known plant species (65), and recent studies have evaluated the roles of the expanded *R. irregularis* proteome (107) and its kinome (81, 108) in this symbiotic relationship. Nevertheless, additional investigation of the contribution of *R. irregularis* pseudokinases in plant-fungal symbiosis is warranted. Furthermore, an understanding of how the symbiotic or infectious nature of host-pathogen interactions operates in the context of bacterial and eukaryotic pseudokinomes is likely to yield important information explaining how such relationships emerged and were propagated during the cellular “wiring” of both physiological and pathological signaling pathways. Hence, the prevalence of pseudokinases in some pathogenic protists such as *P. falciparum* and *G. lamblia* and in some bacteria suggests possible roles in pathogenicity. An understanding of the extent and biological niches for symbiotic and pathogenic pseudokinases will also create further opportunities for pseudokinase-based targeting with small molecules in the future.

Examining the clustering of canonical kinase sequences within pseudokinase families suggests that some pseudokinase families include very closely related, and potentially catalytically active, members (table S3). Although it is currently believed that pseudokinases most likely evolved from gene-duplicated canonical kinases (33), the clustering of canonical sequences within pseudokinase families does not rule out the possibility that some catalytically active kinases might have evolved from “pseudokinases” or other poorly defined ancestral non-enzymatic proteins, as recently reported for other pseudoenzymes (109–111). Some pseudokinase families have quite subtle chemical substitutions at the catalytic triad positions and might therefore be poised to “revert” to a catalytically active enzyme in response to a random mutagenic event or appropriate evolutionary pressure. An artificially guided evolutionary pathway for reversion has been demonstrated for the relatively well-understood pseudokinase CASK, with five steps required to regenerate catalytic activity comparable to a canonical CAMK homolog (112). Likewise, pseudokinase families such as Rig3, HER3, STKLD1, and PSKH2 have conserved all canonical catalytic motifs except for the HRD-aspartate, which is normally substituted to asparagine; canonical members are likely to have evolved in these families by a simple substitution “back” to the canonical HRD-aspartate. Alternatively, kinases can retain low (or very low) amounts of catalytic activity even with aspartate-to-asparagine substitutions in the HRD motif (36, 113, 114), suggesting that pseudokinase families with few relatively benign catalytic motif substitutions may sometimes represent low activity kinases with the capacity for signaling and may explain why canonical sequences sometimes cluster within pseudokinase families. A case for latent catalytic activity can also be made for pseudokinase families such as Act, DYD, and LanC, which conserve the catalytic triad residues except for a benign lysine-to-arginine substitution in the  $\beta$ 3-strand. Nevertheless, the detection of pseudokinase families that have been evolutionarily retained across diverse

species suggests that these predicted pseudokinases cannot be mere “remnants of evolution” but rather that they play important biological roles, either through noncanonical, enzyme-based signaling or, in most of the cases, via noncatalytic functions that await discovery.

Pseudokinases are likely to have evolved because of a relaxed constraint on usually invariant catalytic residues. However, in this study, our examination of the conserved sequence motifs associated with pseudokinase evolutionary divergence reveals variations not only in catalytic motifs but also in conserved “noncatalytic” regions distal from the pseudoactive site (Fig. 7). For example, the observation of pseudokinase family-specific variations at the highly conserved F-helix aspartate suggests that this allosteric region is indeed important for kinase domain activation, as proposed in previous work (103, 115). As a result, future evaluation of pseudokinases may benefit from examining other sequence motifs that contribute to catalysis, including the glycine-rich loop, C-helix glutamate, and the metal-coordinating asparagine, alongside regulatory motifs such as the catalytic and regulatory spines, which comprise allosteric networks distant from the active site. In addition, by comparing two plant IRAK pseudokinase families, we found that inactive kinase domain conformations are stabilized in divergent ways between different pseudokinase families through distinct sets of sequence motifs that have been selectively constrained during evolution. The conservation of pseudokinase family-specific motifs on the surface of the kinase domain further suggests that pseudokinase families have evolved unique interactions with other protein domains or with flexible linker regions. To evidence this, we identified a patch of LRRVI-2 family-specific residues on the surface of the pseudokinase domain that helps tether the C-terminal tail. In terms of regulation, one of the tethering residues is a tyrosine, whose phosphorylation could impart a switch-like function to alter tethering of the C-terminal tail, similar to that observed in the SRC family kinases (116), or for recruiting proteins via the binding of Src homology 2 domains (1, 117). Regulatory tyrosine phosphorylation by dual-specificity LRR RLKs has recently been recognized in plants, where it likely represents a fundamental signaling role despite the lack of conventional TKs encoded in plant kinomes (118–120). The tethering and untethering of flexible flanking linkers is also emerging as a common theme in canonical kinase regulation (55, 121–124) and for modulation of kinase-protein interactions (26, 55, 125, 126). Consistently, these flexible segments are often evolutionary hotspots for neofunctionalization among signaling proteins (116, 127).

This comprehensive curated resource represents a comparative analysis of >30,000 pseudokinase sequences, which includes numerous representative species covering archaea, bacteria, protists, fungi, plants, and animals. It provides a new conceptual framework for characterizing pseudokinase evolution and for the experimental dissection of pseudokinase-dependent lifestyles ranging across a very wide variety of model organisms. Our subclassification of pseudokinases, which includes more than 30 previously unrecognized families, points to fundamental roles for pseudokinases in nearly all biological systems, where reuse of the versatile protein kinase fold has permitted a vast array of noncatalytic signaling mechanisms, many of which might be targeted therapeutically. Our data also represent a useful starting point for the evaluation of other types of pseudoenzymes in diverse biological systems and sets the stage for future evolutionary analyses of other pseudoenzyme families across the kingdoms of life.

**MATERIALS AND METHODS****Detection of pseudokinase sequences**

Protein kinase sequences were extracted from the National Center for Biotechnology Information (NCBI) nonredundant (nr) (downloaded 4 April 2018) and UniProt reference proteome databases (Release 2018\_09) (59). Protein kinase sequences were identified and aligned using previously curated profiles of diverse eukaryotic and eukaryotic-like protein kinases (2, 52, 54, 57, 58) and the rapid and accurate alignment procedure MAPGAPS (56). Some sequences contained many low-complexity regions (for example, in *P. falciparum* and *T. thermophila*), and therefore, a filter was used to mask these low-complexity regions during eukaryotic kinase domain detection. Sequences that did not span from at least the  $\beta$ 3-lysine to the G-helix (like many atypical kinases and small-molecule kinases) were deemed fragmentary and removed.

**Bayesian pattern-based classification of pseudokinase sequences**

We used the pseudokinomes extracted from the UniProt proteomes and additional diverse pseudokinase sequences extracted from the NCBI nr protein database as an input into the omcBPPS (63, 64). To remove closely related sequences, we purged the nr pseudokinase sequences using 75% sequence identity cutoff (22,152 total nr pseudokinase sequences) and restrained the number of UniProt proteomes to 83 diverse representative archaeal, bacterial, and eukaryotic species (4121 total UniProt sequences). A combined total of 26,273 distinct sequences of aligned pseudokinase domains was then used as an input for omcBPPS using a cluster size cutoff of 50 sequences to identify the major sequence families (63, 64). On the basis of this clustering, we initially identified 68 unique pseudokinase clusters. Some human pseudokinases were not classified by the algorithm due to the limited number of sequences (<50 sequences). In these cases, we ran separate clustering (omcBPPS analyses) within these groups using a minimum size of 15 sequences per cluster, which allowed us to further subclassify these clusters. From these subclassifications, we took only the clusters containing human pseudokinases to cover the entire human pseudokinome in our classification, yielding a total of 77 unique pseudokinase clusters in total.

The 77 pseudokinase clusters were incorporated within an existing hierarchical profile of ePK sequences, yielding a total of 592 ePK sequence profiles using MAPGAPS (2, 52, 57, 58). Using the resulting hierarchical sequence profile, we then classified all sequences from each of the 10,092 proteomes available in the UniProt proteome database to all kinase/pseudokinase families in the profile. Pseudokinase sequences that did not classify into the 77 pseudokinase clusters were placed within one of the nine unclassified pseudokinase families based on sequence similarity to the major kinase groups (i.e., TK-, TKL-, STE-, CK1-, AGC-, CAMK-, CMGC-, Other-, and PknB-unclassified groups), resulting in a total of 86 pseudokinase families. For 38 pseudokinase families, canonical kinase sequences clustered into the pseudokinase family. For these cases, canonical sequences were removed from the pseudokinase sequence set and are noted in table S3. Canonical sequences that clustered within the 38 pseudokinase families are provided in data file S3. Pseudokinase family alignments were manually evaluated for possible misalignments, which are noted in table S2.

**Phylogenetic tree building**

To understand the relationships between the identified pseudokinase families, HMMs were built using alignments for each family, and

HMM-to-HMM distances were computed. From these distances, a distance matrix was created to build a neighbor-joining tree, which was used to approximate the distances of families shown in Fig. 3. This method was implemented using pHMM-Tree (128).

The IRAK, *R. irregularis*, and PknB phylogenetic trees were built from aligned kinase domains using FastTree version 2.1.10 using default settings, which implements the JTT ML model and calculates local support values for internal nodes via the Shimodaira-Hasegawa test (129, 130). The rooted IRAK phylogenetic tree was created using a total of 7647 diverse plant and metazoan IRAK sequences. These sequences included all plant IRAK pseudokinase sequences (5405 sequences), canonical plant kinase sequences purged at 80% sequence identity (1894 sequences), metazoan IRAK pseudokinases purged at 90% sequence identity (113 sequences), and canonical metazoan kinases purged at 80% sequence identity (145 sequences). Also included were diverse non-IRAK TKL sequences (41 sequences) and non-TKL kinase sequences (49 sequences), which were used as an outgroup to root the tree. The unrooted *R. irregularis* phylogenetic tree was created using all 787 members of its kinome. The unrooted PknB phylogenetic tree was created using 769 representative pseudokinase sequences assigned to a PknB-related family (removed 376 divergent sequences with conservation log odds score < -10) in addition to a set of representative canonical PknB kinase sequences (511 sequences). iTOL (Interactive Tree of Life) was used to generate the final trees (131).

**WebLogo creation**

WebLogos were created using version 3.6 of the WebLogo 3 online server (132). Amino acids were colored according to their biochemical properties: basic in blue, acidic in red, amide groups in purple, nonpolar residues in black, and polar or uncharged residues in green.

**Determination of domain organizations**

Additional domains for each of the IRAK, *R. irregularis*, and PknB-related pseudokinase families were initially identified using NCBI's Batch Web CD-Search Tool against the CDD database with an expected value threshold of 0.01 and a maximum of 500 hits per CD-search (133, 134). TM helices were identified using TMHMM 2.0 (135) and the Phobius web server (136). TM helices were annotated as such only if both TMHMM and Phobius predicted the same TM helix region within a 10-residue margin. We verified kinase domain hits using our manually curated protein kinase profiles and removed any overlapping domain predictions.

**Pattern analysis of plant IRAK families**

To detect sequence motifs associated with the evolution of IRAK pseudokinase families, we used sequence sets from the omcBPPS classification of IRAK pseudokinase families as seed alignments and used mcBPPS (137) to optimally partition 136,068 diverse IRAK sequences extracted from the NCBI nr database (including active sequences) into either pseudokinase family sets or a "background" set that includes unclassified IRAK sequences. Some pseudokinase families have very close active homologs, and thus, separate seed alignments for canonical IRAK families were also included to ensure that pseudokinase partitions did not include active members. mcBPPS identifies the amino acid patterns that most distinguishes each pseudokinase partition from other sequences, and the 30 most statistically significant patterns for each family were analyzed using available crystal structures.



## Analysis of fungal proteomes and kinomes

To compare the kinome and proteome sizes across different fungal species, we analyzed a total of 448 fungal proteomes obtained from the Ensembl Fungi database (138). Kinases and pseudokinases were identified using previously curated multiple protein alignment profiles and examination of the  $\beta$ -lysine, HRD-aspartate, and DFG-aspartate positions, as detailed above.

## SUPPLEMENTARY MATERIALS

stke.sciencemag.org/cgi/content/full/12/578/eaav3810/DC1

Fig. S1. Expansions of fungal proteomes and kinomes.

Table S1. Kinase and pseudokinase sequence counts detected in 10,092 archaeal, bacterial, and eukaryotic proteomes.

Table S2. Catalog and annotation of pseudokinase families.

Table S3. Counts of canonical sequences classified into pseudokinase families.

Table S4. Distribution of plant IRAK pseudokinase families across diverse plant species.

Table S5. Known plant IRAK pseudokinases and their classifications.

Data file S1. Alignments of model organism kinomes and pseudokinomes.

Data file S2. Sequences and alignments of pseudokinase families.

Data file S3. Alignments of canonical sequences classified into pseudokinase families.

## REFERENCES AND NOTES

- P. Cohen, The origins of protein phosphorylation. *Nat. Cell Biol.* **4**, E127–E130 (2002).
- N. Kannan, S. S. Taylor, Y. Zhai, J. C. Venter, G. Manning, Structural and functional diversity of the microbial kinome. *PLoS Biol.* **5**, e17 (2007).
- D. R. Knighton, J. H. Zheng, L. F. Ten Eyck, V. A. Ashford, N. H. Xuong, S. S. Taylor, J. M. Sowadski, Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 407–414 (1991).
- D. R. Knighton, J. H. Zheng, L. F. ten Eyck, N. H. Xuong, S. S. Taylor, J. Sowadski, Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 414–420 (1991).
- E. D. Scheeff, J. Eswaran, G. Bunkoczi, S. Knapp, G. Manning, Structure of the pseudokinase VRK3 reveals a degraded catalytic site, a highly conserved kinase fold, and a putative regulatory binding site. *Structure* **17**, 128–138 (2009).
- J. A. Adams, Kinetic and catalytic mechanisms of protein kinases. *Chem. Rev.* **101**, 2271–2290 (2001).
- S. K. Hanks, A. M. Quinn, T. Hunter, The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42–52 (1988).
- S. K. Hanks, T. Hunter, The eukaryotic protein kinase superfamily-kinase catalytic domain structure and classification. *FASEB J.* **9**, 576–596 (1995).
- J. M. Murphy, P. D. Mace, P. A. Eyers, Live and let die: Insights into pseudoenzyme mechanisms from structure. *Curr. Opin. Struct. Biol.* **47**, 95–104 (2017).
- E. D. Scheeff, P. E. Bourne, Structural evolution of the protein kinase-like superfamily. *PLoS Comput. Biol.* **1**, e49 (2005).
- M. Kostich, J. English, V. Madison, F. Gheyas, L. Wang, P. Qiu, J. Greene, T. M. Laz, Human members of the eukaryotic protein kinase family. *Genome Biol.* **3**, RESEARCH0043 (2002).
- L. J. Wilson, A. Linley, D. E. Hammond, F. E. Hood, J. M. Coulson, D. J. MacEwan, S. J. Ross, J. R. Slupsky, P. D. Smith, P. A. Eyers, I. A. Prior, New perspectives, opportunities, and challenges in exploring the human protein kinome. *Cancer Res.* **78**, 15–29 (2018).
- G. Manning, G. D. Plowman, T. Hunter, S. Sudarsanam, Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.* **27**, 514–520 (2002).
- G. Manning, S. L. Young, W. T. Miller, Y. Zhai, The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9674–9679 (2008).
- S. Caenepeel, G. Charydczak, S. Sudarsanam, T. Hunter, G. Manning, The mouse kinome: Discovery and comparative genomics of all mouse protein kinases. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11707–11712 (2004).
- G. Manning, D. S. Reiner, T. Lauwaet, M. Dacre, A. Smith, Y. Zhai, S. Svard, F. D. Gillin, The minimal kinome of *Giardia lamblia* illuminates early kinase evolution and unique parasite biology. *Genome Biol.* **12**, R66 (2011).
- M. Zuluski, G. Schulze, R. Braginetz, S. Hartmann, W. X. Schulze, The Arabidopsis Kinome: Phylogeny and evolutionary insights into functional diversification. *BMC Genomics* **15**, 548 (2014).
- G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
- J. Boudeau, D. Miranda-Saavedra, G. J. Barton, D. R. Alessi, Emerging roles of pseudokinases. *Trends Cell Biol.* **16**, 443–452 (2006).
- P. A. Eyers, J. M. Murphy, Dawn of the dead: Protein pseudokinases signal new adventures in cell biology. *Biochem. Soc. Trans.* **41**, 969–974 (2013).
- V. Reiterer, P. A. Eyers, H. Farhan, Day of the dead: Pseudokinases and pseudophosphatases in physiology and disease. *Trends Cell Biol.* **24**, 489–505 (2014).
- J. M. Murphy, Q. Zhang, S. N. Young, M. L. Reese, F. P. Bailey, P. A. Eyers, D. Ungureanu, H. Hammaren, O. Silvennoinen, L. N. Varghese, K. Chen, A. Tripaydonis, N. Jura, K. Fukuda, J. Qin, Z. Nimchuk, M. B. Mudgett, S. Elowe, C. L. Gee, L. Liu, R. J. Daly, G. Manning, J. J. Babon, I. S. Lucet, A robust methodology to subclassify pseudokinases based on their nucleotide-binding properties. *Biochem. J.* **457**, 323–334 (2014).
- J. M. Mendrola, F. Shi, J. H. Park, M. A. Lemmon, Receptor tyrosine kinases with intracellular pseudokinase domains. *Biochem. Soc. Trans.* **41**, 1029–1036 (2013).
- N. S. Dhawan, A. P. Scopton, A. C. Dar, Small molecule stabilization of the KSR inactive state antagonizes oncogenic Ras signalling. *Nature* **537**, 112–116 (2016).
- J. J. Babon, I. S. Lucet, J. M. Murphy, N. A. Nicola, L. N. Varghese, The molecular regulation of Janus kinase (JAK) activation. *Biochem. J.* **462**, 1–13 (2014).
- P. A. Eyers, K. Keeshan, N. Kannan, Tribbles in the 21st century: The evolving roles of tribbles pseudokinases in biology and disease. *Trends Cell Biol.* **27**, 284–298 (2017).
- E. Zeqiraj, D. M. F. van Aalten, Pseudokinases—remnants of evolution or key allosteric regulators? *Curr. Opin. Struct. Biol.* **20**, 772–781 (2010).
- M. Sierla, H. Hörak, K. Overmyer, C. Waszczak, D. Yarmolinsky, T. Maierhofer, J. P. Vainonen, J. Salojärvi, K. Denessiouk, K. Laanemets, K. Töldsepp, T. Vahisalu, A. Gauthier, T. Puukko, L. Paulin, P. Auvinen, D. Geiger, R. Hedrich, H. Kollist, J. Kangasjärvi, The receptor-like pseudokinase GHR1 is required for stomatal closure. *Plant Cell* **30**, 2813–2837 (2018).
- H. Zhang, Q. Zhu, J. Cui, Y. Wang, M. J. Chen, X. Guo, V. S. Tagliabracchi, J. E. Dixon, J. Xiao, Structure and evolution of the Fam20 kinases. *Nat. Commun.* **9**, 1218 (2018).
- C. Lecointre, V. Simon, C. Kerneur, F. Allemand, A. Fournet, I. Montarras, J. L. Pons, M. Gelin, C. Brignatz, S. Urbach, G. Labesse, S. Roche, Dimerization of the Pragmin pseudo-kinase regulates protein tyrosine phosphorylation. *Structure* **26**, 545–554.e4 (2018).
- W.-Y. Chen, Y.-C. Tsai, H.-L. Yeh, F. Suau, K.-C. Jiang, A.-N. Shao, J. Huang, Y.-N. Liu, Loss of SPDEF and gain of TGFBI activity after androgen deprivation therapy promote EMT and bone metastasis of prostate cancer. *Sci. Signal.* **10**, eaam6826 (2017).
- M. Zettl, C. Adrain, K. Strisovsky, V. Lastun, M. Freeman, Rhomboid family pseudoproteases use the ER quality control machinery to regulate intercellular signaling. *Cell* **145**, 79–91 (2011).
- P. A. Eyers, J. M. Murphy, The evolving world of pseudoenzymes: Proteins, prejudice and zombies. *BMC Biol.* **14**, 98 (2016).
- J. M. Murphy, H. Farhan, P. A. Eyers, Bio-zombie: The rise of pseudoenzymes in biology. *Biochem. Soc. Trans.* **45**, 537–544 (2017).
- C. J. Novotny, S. Pollari, J. H. Park, M. A. Lemmon, W. Shen, K. M. Shokat, Overcoming resistance to HER2 inhibitors through state-specific kinase binding. *Nat. Chem. Biol.* **12**, 923–930 (2016).
- F. Shi, S. E. Telesco, Y. Liu, R. Radhakrishnan, M. A. Lemmon, ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 7692–7697 (2010).
- B. Xu, J. M. English, J. L. Wilsbacher, S. Stippec, E. J. Goldsmith, M. H. Cobb, WNK1, a novel mammalian serine/threonine protein kinase lacking the catalytic lysine in subdomain II. *J. Biol. Chem.* **275**, 16795–16801 (2000).
- A.-Y. Lee, W. Chen, S. Stippec, J. Self, F. Yang, X. Ding, S. Chen, Y.-C. Juang, M. H. Cobb, Protein kinase WNK3 regulates the neuronal splicing factor Fox-1. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16841–16846 (2012).
- R. Ahlstrom, A. S. L. Yu, Characterization of the kinase activity of a WNK4 protein complex. *Am. J. Physiol. Renal Physiol.* **297**, F685–F692 (2009).
- P. A. Eyers, Tribbles: A twist in the pseudokinase tail. *Structure* **23**, 1974–1976 (2015).
- S. Uljon, X. Xu, I. Durzynska, S. Stein, G. Adelmant, J. A. Marto, W. S. Pear, S. C. Blacklow, Structural basis for substrate selectivity of the E3 ligase COP1. *Structure* **24**, 687–696 (2016).
- S. A. Jamieson, Z. Ruan, A. E. Burgess, J. R. Curry, H. D. McMillan, J. L. Brewster, A. K. Dunbier, A. D. Axtman, N. Kannan, P. D. Mace, Substrate binding allosterically relieves autoinhibition of the pseudokinase TRB1. *Sci. Signal.* **11**, eaau0597 (2018).
- J. B. Sheetz, M. A. Lemmon, Flipping ATP to amplify kinase functions. *Cell* **175**, 641–642 (2018).
- A. Sreelatha, S. S. Yee, V. A. Lopez, B. C. Park, L. N. Kinch, S. Pilch, K. A. Servage, J. Zhang, J. Jiou, M. Karasiewicz-Urbańska, M. Łobocka, N. V. Grishin, K. Orth, R. Kucharczyk, K. Pawłowski, D. R. Tomchick, V. S. Tagliabracchi, Protein AMPylation by an evolutionarily conserved pseudokinase. *Cell* **175**, 809–821.e19 (2018).
- N. V. Sergina, M. Rausch, D. Wang, J. Blair, B. Hann, K. M. Shokat, M. M. Moasser, Escape from HER-family tyrosine kinase inhibitor therapy by the kinase-inactive HER3. *Nature* **445**, 437–441 (2007).
- J. Claus, G. Patel, F. Autore, A. Colomba, G. Weitsman, T. N. Soliman, S. Roberts, L. C. Zanetti-Domingues, M. Hirsch, F. Collu, R. George, E. Ortiz-Zapater, P. R. Barber,



- B. Vojnovic, Y. Yarden, M. L. Martin-Fernandez, A. Cameron, F. Fraternali, T. Ng, P. J. Parker, Inhibitor-induced HER2-HER3 heterodimerisation promotes proliferation through a novel dimer interface. *eLife* **7**, e32271 (2018).
47. F. P. Bailey, D. P. Byrne, K. Oruganty, C. E. Evers, C. J. Novotny, K. M. Shokat, N. Kannan, P. A. Evers, The Tribbles 2 (TRB2) pseudokinase binds to ATP and autophosphorylates in a metal-independent manner. *Biochem. J.* **467**, 47–62 (2015).
48. G. Labesse, M. Gelin, Y. Bessin, M. Lebrun, J. Papoin, R. Cerdan, S. T. Arold, J. F. Dubremetz, ROP2 from *Toxoplasma gondii*: A virulence factor with a protein-kinase fold and no enzymatic activity. *Structure* **17**, 139–146 (2009).
49. B. S. Blaum, S. Mazzotta, E. R. Nöldeke, T. Halter, J. Madlung, B. Kemmerling, T. Stehle, Structure of the pseudokinase domain of BIR2, a regulator of BAK1-mediated immune signaling in *Arabidopsis*. *J. Struct. Biol.* **186**, 112–121 (2014).
50. L. A. Gish, S. E. Clark, The RLK/Pelle family of kinases. *Plant J.* **66**, 117–127 (2011).
51. J. D. Lewis, A. H.-Y. Lee, J. A. Hassan, J. Wan, B. Hurley, J. R. Jhingree, P. W. Wang, T. Lo, J.-Y. Youn, D. S. Guttman, D. Desveaux, The *Arabidopsis* ZED1 pseudokinase is required for ZAR1-mediated immunity induced by the *Pseudomonas syringae* type III effector HopZ1a. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18722–18727 (2013).
52. M. D. Lehti-Shiu, S.-H. Shiu, Diversity, classification and function of the plant protein kinase superfamily. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **367**, 2619–2639 (2012).
53. P.-L. Liu, L. Du, Y. Huang, S.-M. Gao, M. Yu, Origin and diversification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in plants. *BMC Evol. Biol.* **17**, 47 (2017).
54. A. Bateman, L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. Sonnhammer, D. J. Studholme, C. Yeats, S. R. Eddy, The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141 (2004).
55. N. Kannan, N. Haste, S. S. Taylor, A. F. Neuwald, The hallmark of AGC kinase functional divergence is its C-terminal tail, a cis-acting regulatory module. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1272–1277 (2007).
56. A. F. Neuwald, Rapid detection, classification and accurate alignment of up to a million or more related protein sequences. *Bioinformatics* **25**, 1869–1875 (2009).
57. D. I. McSkimming, S. Dastgheib, T. R. Baffi, D. P. Byrne, S. Ferries, S. T. Scott, A. C. Newton, C. E. Evers, K. J. Kochut, P. A. Evers, N. Kannan, KinView: A visual comparative sequence analysis tool for integrated kinome research. *Mol. BioSyst.* **12**, 3651–3665 (2016).
58. E. Talevich, A. Mirza, N. Kannan, Structural and evolutionary divergence of eukaryotic protein kinases in Apicomplexa. *BMC Evol. Biol.* **11**, 321 (2011).
59. The UniProt Consortium, UniProt: A hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
60. O. Patel, M. D. W. Griffin, S. Panjikar, W. Dai, X. Ma, H. Chan, C. Zheng, A. Kropp, J. M. Murphy, R. J. Daly, I. S. Lucet, Structure of SgK223 pseudokinase reveals novel mechanisms of homotypic and heterotypic association. *Nat. Commun.* **8**, 1157 (2017).
61. S. B. McMahon, H. A. van Buskirk, K. A. Dugan, T. D. Copeland, M. D. Cole, The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. *Cell* **94**, 363–374 (1998).
62. P. Ward, L. Equinet, J. Packer, C. Doerig, Protein kinases of the human malaria parasite *Plasmodium falciparum*: The kinome of a divergent eukaryote. *BMC Genomics* **5**, 79 (2004).
63. A. F. Neuwald, A Bayesian sampler for optimization of protein domain hierarchies. *J. Comput. Biol.* **21**, 269–286 (2014).
64. A. F. Neuwald, Evaluating, comparing, and interpreting protein domain hierarchies. *J. Comput. Biol.* **21**, 287–302 (2014).
65. E. Tisserant, M. Malbreil, A. Kuo, A. Kohler, A. Symeonidi, R. Balestrini, P. Charron, N. Duensing, N. Frei di Frey, V. Gianinazzi-Pearson, L. B. Gilbert, Y. Handa, J. R. Herr, M. Hijiri, R. Koul, M. Kawaguchi, F. Krajinski, P. J. Lammers, F. G. Masclaux, C. Murat, E. Morin, S. Ndikumana, M. Pagni, D. Petitpierre, N. Requena, P. Roskiewicz, R. Riley, K. Saito, H. S. Clemente, H. Shapiro, D. van Tuinen, G. Bécard, P. Bonfante, U. Paszkowski, Y. Y. Shachar-Hill, G. A. Tuskan, J. P. W. Young, I. R. Sanders, B. Henrissat, S. A. Rensing, I. V. Grigoriev, N. Corradi, C. Roux, F. Martin, Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20117–20122 (2013).
66. S. L. Sierke, K. Cheng, H. H. Kim, J. G. Koland, Biochemical characterization of the protein tyrosine kinase homology domain of the ErbB3 (HER3) receptor protein. *Biochem. J.* **322**, 757–763 (1997).
67. A. Miya, P. Albert, T. Shinya, Y. Desaki, K. Ichimura, K. Shirasu, Y. Narusaka, N. Kawakami, H. Kaku, N. Shibuya, CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19613–19618 (2007).
68. S. Radutoiu, L. H. Madsen, E. B. Madsen, H. H. Felle, Y. Umehara, M. Grønlund, S. Sato, Y. Nakamura, S. Tabata, N. Sandal, J. Stougaard, Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**, 585–592 (2003).
69. A. Pietraszewska-Bogiel, B. Lefebvre, M. A. Koini, D. Klaus-Heisen, F. L. W. Takken, R. Geurts, J. V. Cullimore, T. W. J. Gadella, Interaction of *Medicago truncatula* lysin motif receptor-like kinases, NFP and LYK3, produced in *Nicotiana benthamiana* induces defence-like responses. *PLoS ONE* **8**, e65055 (2013).
70. B. Eyuboglu, K. Pfister, G. Haberer, D. Chevalier, A. Fuchs, K. F. X. Mayer, K. Schneitz, Molecular characterisation of the STRUBBELIG-RECEPTOR FAMILY of genes encoding putative leucine-rich repeat receptor-like kinases in *Arabidopsis thaliana*. *BMC Plant Biol.* **7**, 16 (2007).
71. D. Chevalier, M. Batoux, L. Fulton, K. Pfister, R. K. Yadav, M. Schellenberg, K. Schneitz, STRUBBELIG defines a receptor kinase-mediated signaling pathway regulating organ development in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9074–9079 (2005).
72. T. Miyakawa, K.-i. Miyazono, Y. Sawano, K.-i. Hatano, M. Tanokura, Crystal structure of ginkgobilin-2 with homology to the extracellular domain of plant cysteine-rich receptor-like kinases. *Proteins* **77**, 247–251 (2009).
73. L. Zhang, L.-H. Tian, J.-F. Zhao, Y. Song, C.-J. Zhang, Y. Guo, Identification of an apoplastic protein involved in the initial phase of salt stress response in rice root by two-dimensional electrophoresis. *Plant Physiol.* **149**, 916–928 (2009).
74. G. Bourdais, P. Burdiak, A. Gauthier, L. Nitsch, J. Salojärvi, C. Rayapuram, N. Idänheimo, K. Hunter, S. Kimura, E. Merilo, A. Vaattovaara, K. Oracz, D. Kaufholdt, A. Pallon, D. T. Anggoro, D. Glów, J. Lowe, J. Zhou, O. Mohammadi, T. Puukko, A. Albert, H. Lang, D. Ernst, H. Kollist, M. Brosché, J. Durner, J. W. Borst, D. B. Collinge, S. Karpiński, M. F. Lyngkjær, S. Robatzek, M. Wrzaczek, J. Kangasjärvi; CRK Consortium, Large-scale phenomics identifies primary and fine-tuning roles for CRKs in responses related to oxidative stress. *PLoS Genet.* **11**, e1005373 (2015).
75. K. Sklodowski, J. Riedelsberger, N. Raddatz, G. Riadi, J. Caballero, I. Chérel, W. Schulze, A. Graf, I. Dreyer, The receptor-like pseudokinase MRH1 interacts with the voltage-gated potassium channel AKT2. *Sci. Rep.* **7**, 44611 (2017).
76. C. Huard-Chauveau, L. Perchepped, M. Debieu, S. Rivas, T. Kroj, I. Kars, J. Bergelson, F. Roux, D. Roby, An atypical kinase under balancing selection confers broad-spectrum disease resistance in *Arabidopsis*. *PLoS Genet.* **9**, e1003766 (2013).
77. C. Grütter, S. Sreeramulu, G. Sessa, D. Rauh, Structural characterization of the RLCK family member BSK8: A pseudokinase with an unprecedented architecture. *J. Mol. Biol.* **425**, 4455–4467 (2013).
78. M. Gao, X. Wang, D. Wang, F. Xu, X. Ding, Z. Zhang, D. Bi, Y. T. Cheng, S. Chen, X. Li, Y. Zhang, Regulation of cell death and innate immunity by two receptor-like kinases in *Arabidopsis*. *Cell Host Microbe* **6**, 34–44 (2009).
79. Z. L. Nimchuk, P. T. Tarr, E. M. Meyerowitz, An evolutionarily conserved pseudokinase mediates stem cell production in plants. *Plant Cell* **23**, 851–854 (2011).
80. A. Kuo, A. Kohler, F. M. Martin, I. V. Grigoriev, Expanding genomics of mycorrhizal symbiosis. *Front. Microbiol.* **5**, 582 (2014).
81. Y. Handa, H. Nishide, N. Takeda, Y. Suzuki, M. Kawaguchi, K. Saito, RNA-seq transcriptional profiling of an arbuscular mycorrhiza provides insights into regulated and coordinated gene expression in *Lotus japonicus* and *Rhizophagus irregularis*. *Plant Cell Physiol.* **56**, 1490–1511 (2015).
82. C. E. Haydon, P. A. Evers, L. D. Aveline-Wolf, K. A. Resing, J. L. Maller, N. G. Ahn, Identification of novel phosphorylation sites on *Xenopus laevis* Aurora A and analysis of phosphopeptide enrichment by immobilized metal-affinity chromatography. *Mol. Cell. Proteomics* **2**, 1055–1067 (2003).
83. M. Grynberg, A. Godzik, NERD: A DNA processing-related domain present in the anthrax virulence plasmid, pXO1. *Trends Biochem. Sci.* **29**, 106–110 (2004).
84. P. S. Sumbly, M. C. M. Smith, Genetics of the phage growth limitation (Pgl) system of *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **44**, 489–500 (2002).
85. P. A. Hoskisson, P. Sumbly, M. C. M. Smith, The phage growth limitation system in *Streptomyces coelicolor* A(3)2 is a toxin/antitoxin system, comprising enzymes with DNA methyltransferase, protein kinase and ATPase activity. *Virology* **477**, 100–109 (2015).
86. R. H. Ebricht, S. Busby, The *Escherichia coli* RNA polymerase alpha subunit: Structure and function. *Curr. Opin. Genet. Dev.* **5**, 197–203 (1995).
87. K. Aoyama, H. Aiba, T. Mizuno, Genetic analysis of the His-to-Asp phosphorelay implicated in mitotic cell cycle control: Involvement of histidine-kinase genes of *Schizosaccharomyces pombe*. *Biosci. Biotechnol. Biochem.* **65**, 2347–2352 (2001).
88. M. K. Ashby, J. Houmar, Cyanobacterial two-component proteins: Structure, diversity, distribution, and evolution. *Microbiol. Mol. Biol. Rev.* **70**, 472–509 (2006).
89. J. A. Calera, G. H. Choi, R. A. Calderone, Identification of a putative histidine kinase two-component phosphorelay gene (CaHK1) in *Candida albicans*. *Yeast* **14**, 665–674 (1998).
90. A. Kupra, N. Srinivasan, Diversity in domain architectures of Ser/Thr kinases and their homologues in prokaryotes. *BMC Genomics* **6**, 129 (2005).
91. V. Phalip, J. H. Li, C. C. Zhang, HistK a cyanobacterial protein with both a serine/threonine kinase domain and a histidine kinase domain: Implication for the mechanism of signal transduction. *Biochem. J.* **360**, 639–644 (2001).
92. X. Zhang, F. Zhao, X. Guan, Y. Yang, C. Liang, S. Qin, Genome-wide survey of putative serine/threonine protein kinases in cyanobacteria. *BMC Genomics* **8**, 395 (2007).
93. V. Buck, J. Quinn, T. S. Pino, H. Martin, J. Saldanha, K. Makino, B. A. Morgan, J. B. A. Millar, Peroxide sensors for the fission yeast stress-activated mitogen-activated protein kinase pathway. *Mol. Biol. Cell* **12**, 407–419 (2001).
94. J. A. Calera, X.-J. Zhao, F. De Bernardis, M. Sheridan, R. Calderone, Avirulence of *Candida albicans* CaHK1 mutants in a murine model of hematogenously disseminated candidiasis. *Infect. Immun.* **67**, 4280–4284 (1999).

95. J. A. Calera, R. Calderone, Flocculation of hyphae is associated with a deletion in the putative CaHK1 two-component histidine kinase gene from *Candida albicans*. *Microbiology* **145**, 1431–1442 (1999).
96. Y. Cheng, J.-H. Li, L. Shi, L. Wang, A. Latifi, C.-C. Zhang, A pair of iron-responsive genes encoding protein kinases with a Ser/Thr kinase domain and a His kinase domain are regulated by NtcA in the Cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **188**, 4822–4829 (2006).
97. M. Kruppa, T. Goins, J. E. Cutler, D. Lowman, D. Williams, N. Chauhan, V. Menon, P. Singh, D. Li, R. Calderone, The role of the *Candida albicans* histidine kinase (*CHK1*) gene in the regulation of cell wall mannan and glucan biosynthesis. *FEMS Yeast Res.* **3**, 289–299 (2003).
98. J. Quinn, P. Malakasi, D. A. Smith, J. Cheetham, V. Buck, J. B. A. Millar, B. A. Morgan, Two-component mediated peroxide sensing and signal transduction in fission yeast. *Antioxid. Redox Signal.* **15**, 153–165 (2011).
99. L. Shi, J.-H. Li, Y. Cheng, L. Wang, W.-L. Chen, C.-C. Zhang, Two genes encoding protein kinases of the HstK family are involved in synthesis of the minor heterocyst-specific glycolipid in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **189**, 5075–5081 (2007).
100. A. Torosantucci, P. Chiani, F. De Bernardis, A. Cassone, J. A. Calera, R. Calderone, Deletion of the two-component histidine kinase gene (*CHK1*) of *Candida albicans* contributes to enhanced growth inhibition and killing by human neutrophils in vitro. *Infect. Immun.* **70**, 985–987 (2002).
101. T. Yamada-Okabe, T. Mio, N. Ono, Y. Kashima, M. Matsui, M. Arisawa, H. Yamada-Okabe, Roles of three histidine kinase genes in hyphal development and virulence of the pathogenic fungus *Candida albicans*. *J. Bacteriol.* **181**, 7243–7247 (1999).
102. M. A. Jones, M. J. Raymond, N. Smirnov, Analysis of the root-hair morphogenesis transcriptome reveals the molecular identity of six genes with roles in root-hair development in *Arabidopsis*. *Plant J.* **45**, 83–100 (2006).
103. K. Oruganty, N. S. Talathi, Z. A. Wood, N. Kannan, Identification of a hidden strain switch provides clues to an ancient structural mechanism in protein kinases. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 924–929 (2013).
104. T. Hunter, G. D. Plowman, The protein kinases of budding yeast: Six score and more. *Trends Biochem. Sci.* **22**, 18–22 (1997).
105. J. E. Stajich, S. K. Wilke, D. Ahrén, C. H. Au, B. W. Birren, M. Borodovsky, C. Burns, B. Canbäck, L. A. Casselton, C. K. Cheng, J. Deng, F. S. Dietrich, D. C. Fargo, M. L. Farman, A. C. Gathman, J. Goldberg, R. Guigó, P. J. Hoegger, J. B. Hooker, A. Huggins, T. Y. James, T. Kamada, S. Kilaru, C. Kodira, U. Kües, D. Kupfer, H. S. Kwan, A. Lomsadze, W. Li, W. W. Lilly, L.-J. Ma, A. J. Mackey, G. Manning, F. Martin, H. Muraguchi, D. O. Natvig, H. Palmerini, M. A. Ramesh, C. J. Rehmeyer, B. A. Roe, N. Shenoy, M. Stanke, V. Ter-Hovhannisyan, A. Tunlid, R. Velagapudi, T. J. Vision, Q. Zeng, M. E. Zolan, P. J. Pukkila, Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11889–11894 (2010).
106. W. T. Miller, Tyrosine kinase signaling and the emergence of multicellularity. *Biochim. Biophys. Acta* **1823**, 1053–1057 (2012).
107. G. D. A. Werner, Y. Zhou, C. M. J. Pieterse, E. T. Kiers, Tracking plant preference for higher-quality mycorrhizal symbionts under varying CO<sub>2</sub> conditions over multiple generations. *Ecol. Evol.* **8**, 78–87 (2018).
108. K. Nanjareddy, M. K. Arthikala, B. M. Gómez, L. Blanco, M. Lara, Differentially expressed genes in mycorrhizal and nodulated roots of common bean are associated with defense, cell wall architecture, N metabolism, and P metabolism. *PLOS ONE* **12**, e0182328 (2017).
109. A. E. Todd, C. A. Orenge, J. M. Thornton, Sequence and structural differences between enzyme and nonenzyme homologs. *Structure* **10**, 1435–1451 (2002).
110. M. Kaltenbach, J. R. Burke, M. Dindo, A. Pabis, F. S. Munsberg, A. Rabin, S. C. L. Kamerlin, J. P. Noel, D. S. Tawfik, Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat. Chem. Biol.* **14**, 548–555 (2018).
111. B. E. Clifton, J. A. Kaczmarek, P. D. Carr, M. L. Gerth, N. Tokuriki, C. J. Jackson, Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. *Nat. Chem. Biol.* **14**, 542–547 (2018).
112. K. Mukherjee, M. Sharma, R. Jahn, M. C. Wahl, T. C. Sudhof, Evolution of CASK into a Mg<sup>2+</sup>-sensitive kinase. *Sci. Signal.* **3**, ra33 (2010).
113. D. M. Williams, P. A. Cole, Proton demand inversion in a mutant protein tyrosine kinase reaction. *J. Am. Chem. Soc.* **124**, 5956–5957 (2002).
114. V. T. Skamnaki, D. J. Owen, M. E. M. Noble, E. D. Lowe, G. Lowe, N. G. Oikonomakos, L. N. Johnson, Catalytic mechanism of phosphorylase kinase probed by mutational studies. *Biochemistry* **38**, 14718–14730 (1999).
115. S. Mohanty, K. Oruganty, A. Kwon, D. P. Byrne, S. Ferries, Z. Ruan, L. E. Hanold, S. Katiyar, E. J. Kennedy, P. A. Eyers, N. Kannan, Hydrophobic core variations provide a structural framework for tyrosine kinase evolution and functional specialization. *PLOS Genet.* **12**, e1005885 (2016).
116. S. C. Harrison, Variation on an Src-like theme. *Cell* **112**, 737–740 (2003).
117. J. G. Williams, M. Zvelebil, SH2 domains in plants imply new signalling scenarios. *Trends Plant Sci.* **9**, 161–163 (2004).
118. M.-H. Oh, X. Wang, U. Kota, M. B. Goshe, S. D. Clouse, S. C. Huber, Tyrosine phosphorylation of the BRI1 receptor kinase emerges as a component of brassinosteroid signaling in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 658–663 (2009).
119. S. Luan, Tyrosine phosphorylation in plant cell signaling. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11567–11569 (2002).
120. A. Perraki, T. A. DeFalco, P. Derbyshire, J. Avila, D. Sére, J. Sklenar, X. Qi, L. Stransfeld, B. Schwesinger, Y. Kadota, A. P. Macho, S. Jiang, D. Couto, K. U. Torii, F. L. H. Menke, C. Zipfel, Phosphocode-dependent functional dichotomy of a common co-receptor in plant signalling. *Nature* **561**, 248–252 (2018).
121. S. R. Hubbard, Juxtamembrane autoinhibition in receptor tyrosine kinases. *Nat. Rev. Mol. Cell Biol.* **5**, 464–471 (2004).
122. A. Kwon, M. John, Z. Ruan, N. Kannan, Coupled regulation by the juxtamembrane and sterile alpha motif (SAM) linker is a hallmark of ephrin tyrosine kinase evolution. *J. Biol. Chem.* **293**, 5102–5116 (2018).
123. A. Mirza, M. Mustafa, E. Talevich, N. Kannan, Co-conserved features associated with cis regulation of ErbB tyrosine kinases. *PLOS ONE* **5**, e14310 (2010).
124. I. Plaza-Menacho, K. Barnouin, R. Barry, A. Borg, M. Orme, R. Chauhan, S. Mouilleron, R. J. Martínez-Torres, P. Meier, N. Q. McDonald, RET functions as a dual-specificity kinase that requires allosteric inputs from Juxtamembrane elements. *Cell Rep.* **17**, 3319–3332 (2016).
125. K. S. Gajiwala, EGFR: Tale of the C-terminal tail. *Protein Sci.* **22**, 995–999 (2013).
126. B. Ma, C.-J. Tsai, T. Haliloglu, R. Nussinov, Dynamic allostery: Linkers are not merely flexible. *Structure* **19**, 907–917 (2011).
127. C. J. Brown, A. K. Johnson, A. K. Dunker, G. W. Daughdrill, Evolution and disorder. *Curr. Opin. Struct. Biol.* **21**, 441–446 (2011).
128. L. Huo, H. Zhang, X. Huo, Y. Yang, X. Li, Y. Yin, pHMM-tree: Phylogeny of profile hidden Markov models. *Bioinformatics* **33**, 1093–1095 (2017).
129. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
130. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010).
131. I. Letunic, P. Bork, Interactive tree of life (ITOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
132. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
133. A. Marchler-Bauer, Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, L. Y. Geer, S. H. Bryant, CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45**, D200–D203 (2017).
134. A. Marchler-Bauer, S. Lu, J. B. Anderson, F. Chitsaz, M. K. Derbyshire, C. DeWeese-Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, F. Lu, G. H. Marchler, M. Mullokkandov, M. V. Omelchenko, C. L. Robertson, J. S. Song, N. Thanki, R. A. Yamashita, D. Zhang, N. Zhang, C. Zheng, S. H. Bryant, CDD: A conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225–D229 (2011).
135. A. Krogh, B. Larsson, G. von Heijne, E. L. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
136. L. Kall, A. Krogh, E. L. L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction—The Phobius web server. *Nucleic Acids Res.* **35**, W429–W432 (2007).
137. A. F. Neuwald, Surveying the manifold divergence of an entire protein class for statistical clues to underlying biochemical mechanisms. *Stat. Appl. Genet. Mol. Biol.* **10**, 36 (2011).
138. P. J. Kersey, J. E. Allen, A. Allot, M. Barba, S. Boddu, B. J. Bolt, D. Carvalho-Silva, M. Christensen, P. Davis, C. Grabmueller, N. Kumar, Z. Liu, T. Maurel, B. Moore, M. D. McDowall, U. Maheswari, G. Naamati, V. Newman, C. K. Ong, M. Paulini, H. Pedro, E. Perry, M. Russell, H. Sparrow, E. Tapanari, K. Taylor, A. Vullo, G. Williams, A. Zaddissia, A. Olson, J. Stein, S. Wei, M. Tello-Ruiz, D. Ware, A. Luciani, S. Potter, R. D. Finn, M. Urban, K. E. Hammond-Kosack, D. M. Bolser, N. De Silva, K. L. Howe, N. Langridge, G. Maslen, D. M. Staines, A. Yates, Ensembl genomes 2018: An integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res.* **46**, D802–D808 (2018).

**Acknowledgments:** We acknowledge members of the Kannan Laboratory and A. F. Neuwald for helpful discussions. **Funding:** This work was supported by the NIH (R01GM1149 to N.K.), the NSF (MCB-1149106 to N.K.), and a UGA-Liverpool travel fellowship (to A.K.). This work was also funded by a BBSRC Tools and Resources Development Fund award (BB/N021703/1), a Royal Society Research Grant, and North West Cancer Research grants (CR1088 and CR1097 to

P.A.E.). **Author contributions:** A.K., P.A.E., and N.K. conceptualized the work. A.K., S.S., R.T., W.Y., and N.K. curated and analyzed sequence data. A.K., S.S., R.T., W.Y., P.A.E., and N.K. co-wrote the manuscript. K.J.K. integrated sequence data into ProKinO. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Sequence data have been made available through ProKinO (<http://vulcan.cs.uga.edu/prokino/about/browser>). All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials.

Submitted 12 September 2018

Accepted 1 April 2019

Published 23 April 2019

10.1126/scisignal.aav3810

**Citation:** A. Kwon, S. Scott, R. Tadjale, W. Yeung, K. J. Kochut, P. A. Evers, N. Kannan, Tracing the origin and evolution of pseudokinases across the tree of life. *Sci. Signal.* **12**, eaav3810 (2019).

Confidential  
Do not duplicate

## Abstract

**One-sentence summary:** Pseudokinases are prevalent across species and contribute diverse, noncatalytic signaling functions.

**Editor's Summary:**  
**Hunting pseudokinases**

Pseudokinases generally have a similar structure to that of kinases but without the catalytic function. Rather, they serve alternative, some as yet unknown, functions across various signaling niches in many cell and tissue types. Kwon *et al.* performed an evolutionary analysis of pseudokinases across species and classified both known and previously unknown pseudokinases into groups and families. Their analysis reveals that pseudokinases are more prevalent than was previously appreciated and shows how pseudokinase structures may have evolved to confer advantageous functions in bacteria, plants, and animals. The classification is part of a new online, mineable resource called ProKinO.

Confidential  
Do not duplicate