

University of Dundee

The protein kinases of Dictyostelia and their incorporation into a signalome

Kin, Koryu; Chen, Zhi-Hui; Forbes, Gillian; Lawal, Hajara; Schilde, Christina; Singh, Reema

Published in:
Cellular Signalling

DOI:
[10.1016/j.cellsig.2023.110714](https://doi.org/10.1016/j.cellsig.2023.110714)

Publication date:
2023

Licence:
CC BY

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):

Kin, K., Chen, Z-H., Forbes, G., Lawal, H., Schilde, C., Singh, R., Cole, C., Barton, G. J., & Schaap, P. (2023). The protein kinases of Dictyostelia and their incorporation into a signalome. *Cellular Signalling*, 108, [110714]. <https://doi.org/10.1016/j.cellsig.2023.110714>

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



The protein kinases of Dictyostelia and their incorporation into a signalome

Koryu Kin^{a,c}, Zhi-hui Chen^a, Gillian Forbes^{a,d}, Hajara Lawal^a, Christina Schilde^{a,e},
Reema Singh^{a,b,f}, Christian Cole^{b,g}, Geoffrey J. Barton^b, Pauline Schaap^{a,*}

^a Molecular Cell and Developmental Biology, School of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom

^b Computational Biology, School of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom

^c Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Spain

^d Institut de Génomique Fonctionnelle de Lyon (IGFL), CNRS, École Normale Supérieure de Lyon and Université Claude Bernard Lyon-1, Lyon 69007, France

^e D'Arcy Thompson Unit, School of Life Sciences, University of Dundee, DD1 4HN, United Kingdom

^f Vaccine and Infectious Disease Organization, University of Saskatchewan, 120 Veterinary Road, Saskatoon, SK S7N 5E3, Canada

^g Population Health and Genomics, School of Medicine, University of Dundee, Ninewells Hospital, Dundee DD1 9SY, United Kingdom

ARTICLE INFO

Keywords:

Social amoeba
Evolution of signalling
G-protein coupled receptors
Small GTPases
Transcription factors
Transcriptome-based cluster analysis

ABSTRACT

Protein kinases are major regulators of cellular processes, but the roles of most kinases remain unresolved. Dictyostelid social amoebas have been useful in identifying functions for 30% of its kinases in cell migration, cytokinesis, vesicle trafficking, gene regulation and other processes but their upstream regulators and downstream effectors are mostly unknown. Comparative genomics can assist to distinguish between genes involved in deeply conserved core processes and those involved in species-specific innovations, while co-expression of genes as evident from comparative transcriptomics can provide cues to the protein complement of regulatory networks. Genomes and developmental and cell-type specific transcriptomes are available for species that span the 0.5 billion years of evolution of Dictyostelia from their unicellular ancestors. In this work we analysed conservation and change in the abundance, functional domain architecture and developmental regulation of protein kinases across the 4 major taxon groups of Dictyostelia. All data are summarized in annotated phylogenetic trees of the kinase subtypes and accompanied by functional information of all kinases that were experimentally studied. We detected 393 different protein kinase domains across the five studied genomes, of which 212 were fully conserved. Conservation was highest (71%) in the previously defined AGC, CAMK, CK1, CMCG, STE and TKL groups and lowest (26%) in the “other” group of typical protein kinases. This was mostly due to species-specific single gene amplification of “other” kinases. Apart from the AFK and α -kinases, the atypical protein kinases, such as the PIKK and histidine kinases were also almost fully conserved. The phylogeny-wide developmental and cell-type specific expression profiles of the protein kinase genes were combined with profiles from the same transcriptomic experiments for the families of G-protein coupled receptors, small GTPases and their GEFs and GAPs, the transcription factors and for all genes that upon lesion generate a developmental defect. This dataset was subjected to hierarchical clustering to identify clusters of co-expressed genes that potentially act together in a signalling network. The work provides a valuable resource that allows researchers to identify protein kinases and other regulatory proteins that are likely to act as intermediates in a network of interest.

1. Introduction

Protein kinases are major agents of post-translational modification and play crucial regulatory roles in most cellular processes. The misregulation of individual protein kinases is a frequent cause of disease and the quest for novel pharmaceutical agents that regulate kinase activity is therefore an intense field of research [1]. Most kinases contain a

conserved catalytic domain that phosphorylates either a tyrosine or serine/threonine residue within a conserved consensus sequence that is specifically targeted by individual protein kinases. Sequence similarities separate the typical human protein kinases into eight major groups, the AGC, CAMK, CK1, CMGC, RGC, STE, TK, and TKL protein kinases [2,3] and most of these groups can also be recognized in other eukaryotes [4–8]. In addition, there are smaller sets of so-called “atypical” protein

* Corresponding author.

E-mail addresses: koryu.kin@upf.edu (K. Kin), gillian.forbes@ens-lyon.fr (G. Forbes), c.schilde@dundee.ac.uk (C. Schilde), res498@usask.ca (R. Singh), c.cole@dundee.ac.uk (C. Cole), g.j.barton@dundee.ac.uk (G.J. Barton), p.schaap@dundee.ac.uk (P. Schaap).

<https://doi.org/10.1016/j.cellsig.2023.110714>

Received 7 March 2023; Received in revised form 2 May 2023; Accepted 12 May 2023

Available online 13 May 2023

0898-6568/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

kinases with catalytic domains that are either very diverged or completely different from the “typical” group. Relatively large groupings in this set are the α -kinases [9], the histidine kinases [10], and the phosphatidylinositol-like kinases (PIKKs) [11].

Dictyostelid social amoebas evolved multicellularity within the otherwise unicellular Amoebozoa. They feed as single cells but come together when starved to form migrating slugs and fruiting structures containing up to four different cell types. This intriguing life cycle, combined with ease of genetic transformation and experimental accessibility has made *Dictyostelium discoideum* a popular system for studying cellular processes like DNA repair [12], cell migration [13], cytoskeletal remodelling [14], phagocytosis and other forms of vesicle trafficking [15], mechanisms underlying human pathology [16–18] as well as regulation of cell differentiation and morphogenesis during development [19], and evolution of multicellularity [20] and sociality [21,22]. Protein kinases participate in regulating these processes and their specific roles and interactions with other regulators are therefore widely studied.

Upon completion of the *D. discoideum* (*Ddis*) genome, a robust analysis of its 285 protein kinases showed that of the eight typical kinase groups, the TK and RGC kinases were missing, but that the atypical α -kinases, the histidine kinases, PIKKs and smaller groups of atypical kinases were well represented [5]. For 89 of these kinases a biological role has been established, but their upstream regulators or downstream targets are more rarely known.

The ~150 known species of Dictyostelia are subdivided into two branches, each containing two major groups [23,24]. *Ddis* resides in group 4 and well annotated genomes and developmental and cell-type specific transcriptomes are available for *Ddis* and *D. purpureum* (*Dpur*, also in group 4) and for *D. lacteum* (*Dlac*), *P. pallidum* (*Ppal*), and *D. fasciculatum* (*Dfas*), which represent groups 3, 2 and 1 respectively [25–28]. Groups 1–3 consist of species that form relatively small clustered or branched fruiting bodies with maximally two cell types, the stalk cells and spores. Many species in these groups can also encyst as single cells, the strategy by which their unicellular amoebozoan ancestors survive starvation. The group 4 species form large solitary fruiting bodies with two more cell types, the cup and basal disc cells. Their slugs show extensive migration, but as a group they have lost the ability to encyst (Romerero et al. 2013; Schilde et al. 2014).

Phylogeny-wide analysis of conservation and change in the presence, regulation and function of signal transduction proteins enables distinction between core regulatory proteins and gene gain or modification that only occurred in some taxa. Compared to a single organism approach, the phylogeny-wide approach allows inference of a hierarchy in the importance of individual proteins in a network. Additionally, when correlated to phenotypic change during species evolution, the genetic change that cause the phenotypic innovation can be inferred. For these reasons, we previously investigated conservation and change across Dictyostelia in three major families of signal transduction proteins, the transcription factors [29], the G-protein coupled receptors [30] and the small GTPases with their activators, the guanine-nucleotide exchange factors (GEFs) and inactivators, the GTPase activating proteins (GAPs) [31]. Additionally, we performed an evolutionary comparative analysis of all *Ddis* genes that upon lesion yield a developmental defect. This group of developmentally essential genes (DEG) contains many signal molecules, receptors and other intermediates of signalling pathways as well as down-stream effectors [28]. In this work we have analysed conservation and change in protein kinase families across the taxon-group representative species mentioned above, inclusive of changes in their functional domain architecture and in the developmental- and cell type specific expression of their genes.

Since proteins that act together in a network need to be expressed at the same stage and in the same cell type, gene co-expression was put forward as a method to identify putative protein interactions [32,33]. Hierarchical clustering is a convenient method to arrange genes in a cladogram according to similarity in their expression profiles, which can

be juxtaposed to a heatmap of the relative expression levels of the genes [34]. Clusters of genes with obviously similar expression patterns can then be considered to reflect a possible interacting network. Including more gene expression data and replicate experiments will progressively lead to improved resolution and robustness of the clusters.

We previously performed hierarchical clustering on the phylogeny-wide expression data of the GTPases and their GEFs and GAPs [31]. Here 70 out of 197 experimentally determined interactions were also predicted by transcriptome cluster analysis. The protein kinases are the last large family of regulatory proteins to be analysed across the four dictyostelid taxon groups. We saw this as an opportunity to construct an interactome of the combined families of cellular regulators and the developmentally essential genes. Combined with other approaches, this hierarchical clustering of transcriptome data of 1313 mostly conserved signal transduction genes provides opportunities to fully reconstruct the pathways that control both the cellular processes and the developmental program of Dictyostelia.

2. Methods

2.1. Data collection and initial classification

The analysis used the inferred proteomes of the almost fully assembled genomes of *Dictyostelium fasciculatum* (NCBI accession: PRJNA40189), *Polysphondylium pallidum* (PRJNA40191), *D. lacteum* (PRJNA305239) and *D. discoideum* (PRJNA201) that represent the major taxon groups 1, 2, 3 and 4 of Dictyostelia, respectively, and was supplemented with *D. purpureum* (PRJNA30991), a second group 4 genome that is less complete, but well annotated [25–28].

Protein kinases are usually subdivided into the large group of typical eukaryote kinases (ePKs) with related catalytic domains and several groupings of atypical protein kinases with a range of different catalytic domains. The ePKs are broadly classified into eight major groups: AGC, CAMK, CK1, CMGC, RGC, STE, TK, and TKL [3], but also contain “other” kinases that do not conform to either group. We used a multi-level hidden Markov model (HMM) library of the major groups that was created by [35] to scan the five proteomes listed above for the ePK subtypes. Additionally, we performed an InterProScan [36] on all proteomes to obtain a listing of all protein functional domains. From this listing we isolated all proteins with Interpro IDs for the atypical α -kinases (IPR004166), ABC1 kinases (IPR004147), AFK kinases (IPR015275), G11 kinases (IPR018865), histidine kinases (IPR005467), PIKKs (IPR000403) and RIO kinases (IPR000687). BRD, TAF1 and TIF families of atypical protein kinases were identified by BLAST search, using previously identified orthologs as bait.

The first HMM pull-down of ePKs contained many duplicate kinase domains. These domains were removed and the set was again scanned with the most recent HMM library “allPK.hmm” downloaded from the “kinomer” website (www.compbio.dundee.ac.uk/kinomer) using HMMER3 [37]. This resulted in assignment of the “other” kinases to either of the major groupings, albeit with low probability. We resorted to phylogenetics to also correctly order the “other” kinases. All kinase domains were aligned by MAFFT with the “L-INS-i” option [38] and the alignment was trimmed using trimAl with the “gappyout” option [39]. A phylogeny was inferred with IQ-TREE [40] under the molecular evolution model “LG + F + R10”, which was selected with ModelFinder [41]. This phylogeny subdivided the kinases into 11 groupings (Fig. 1). Comparison of the domains contained within each grouping with those assigned to specific HMM profiles indicated that with few exceptions, six groupings entirely consisted of AGC, CAMK, CK1, CMGC, STE and TKL kinase domains, while the other five, named clades A-E, showed a mixture of HMM affiliations.

The smaller families of retrieved atypical protein kinases were individually aligned using ClustalOmega [42]. Pilot phylogenetic trees were inferred using IQ-TREE or MrBayes 3.2 [43].

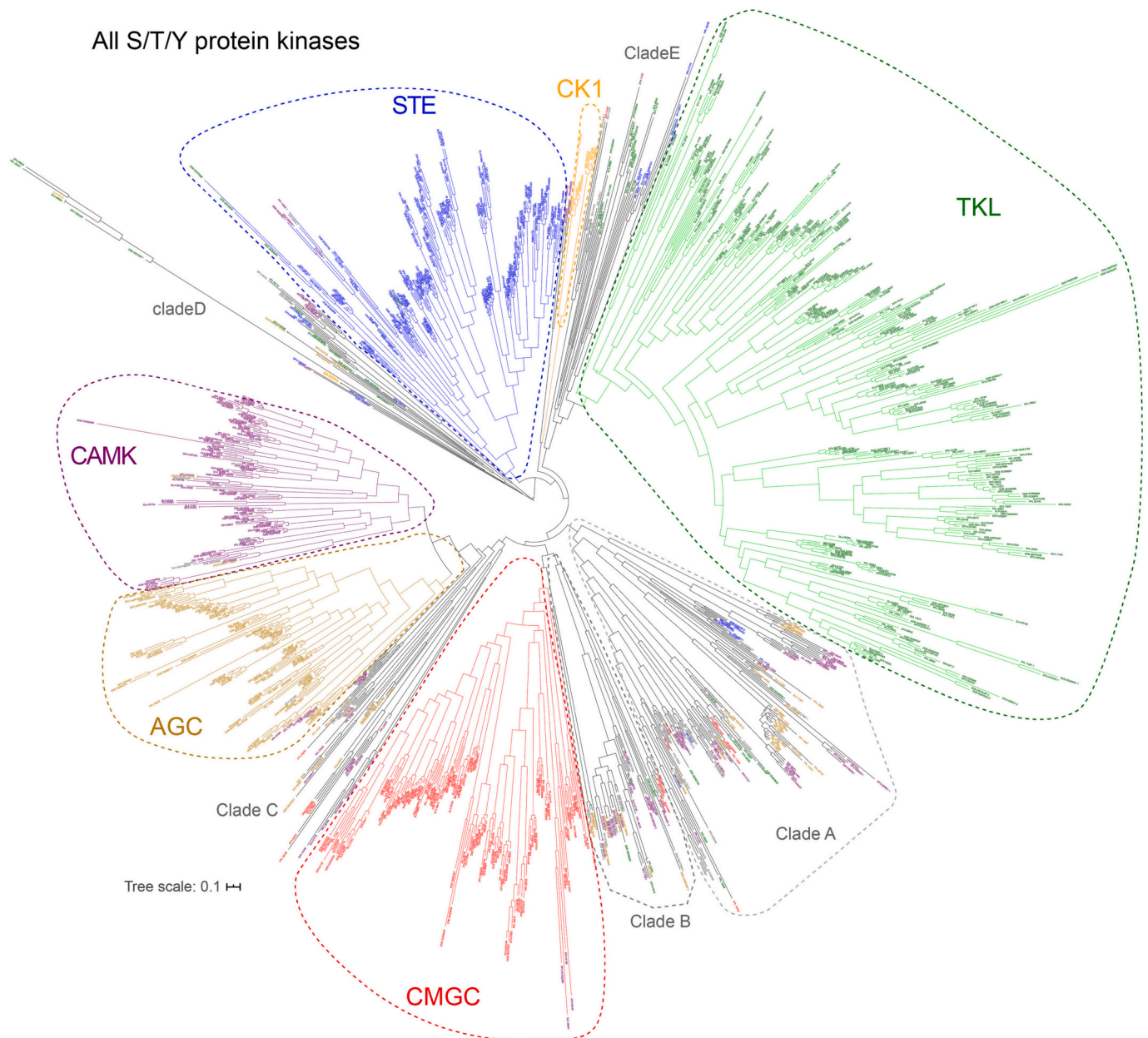


Fig. 1. Phylogeny of “typical” protein kinases of taxon-group representative Dictyostelia.

All “typical” kinase domains identified in the genomes of five representative Dictyostelid species were aligned with MAFFT [38] and a maximum-likelihood phylogeny was inferred with IQ-TREE [40] (see METHODS for details). The phylogenetic tree was visualized on the Interactive Tree Of Life (iTOL) website [77]. The gene names are coloured according to the HMM classification (AGC = brown, CAMK = purple, CK1 = orange, CMGC = red, STE = blue, TKL = green). Six monophyletic clades that consist almost exclusively of genes with the same classification are encircled with dashed lines and labeled accordingly. Five other monophyletic clades are named and labeled as clades A-E. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Curation of kinase phylogenies

The protein sequences of the 6 typical groups and clades A-E of the ePK domains were isolated and aligned separately using MAFFT. Poorly aligned or gappy blocks of sequences were deleted using trimAl. New phylogenies were inferred using either MrBayes 3.2 or IQ-TREE, using mixed amino-acid models and run for 1 million generation or convergence (SD of split frequencies <0.01) for Bayesian inference, or with 1000 bootstrap replicates for IQ-tree. The trees were inspected for missing orthologs in otherwise conserved clades, which were then retrieved by BLASTp or tBLASTn queries of species’ proteomes or genomes, respectively. In this phase we also noted several partial orthologous sets, where the missing likely orthologs were part of another

grouping. The proper classification of the clade was then assessed by realignment and tree inference with either grouping. Once we were confident that all kinases were identified and assigned to their appropriate grouping, their full sets of sequences were aligned and final phylogenies were inferred. The same protocol was followed for the smaller collections of atypical protein kinases.

2.3. Phylogenetic tree annotation

The functional domain architectures, including PFAM domains [44], signal peptides and internal repeats were analysed in the full protein sequences using SMART [45], saved as .svg files and juxtaposed to the protein locus tags at the tips of the phylogenetic tree branches. SMART

or PFAM domain identifiers are listed in each figure and domain descriptions can be retrieved from http://smart.embl-heidelberg.de/smart/domain_table.cgi or <http://pfam.xfam.org/browse>. Clades of orthologous genes or other groupings were annotated with relative transcript levels at specific developmental stages or in specific cell types, shown as heat maps that represent the fraction of the maximum transcript read count for the developmental profiles and the fraction of the summed read counts for the cell types. The normalized reads were retrieved from published RNA sequencing experiments [28,29,46,47].

2.4. Hierarchical clustering

The full set or subsets of relative gene expression data for *Ddis* signal transduction and/or developmentally essential genes and their orthologs in *Dpur*, *Dlac*, *Ppal* and *Dfas*, were ordered into a linear array (Supplemental Table S2 Clusteranalysis.xlsx) and subjected to hierarchical clustering in Orange 3.27.1 [48], using Pearson correlation as the distance metric and average linkage to order the clusters. The data from individual experiments were included for *Ddis* and *Dlac* cell-type specific transcripts, rather than the averaged values used in Fig. 2 and Figs. S1-

S15. Data were standardized as percentage of the maximum value of normalized read counts per gene and per RNAseq experiment in general, or to the sum of normalized read counts, when there were only two data points in the experiment.

3. Results and discussion

3.1. Protein kinase identification and annotation across five dictyostelid genomes

Upon completion of the *D. discoideum* genome [25], its complement of putative protein kinases was rigorously investigated using both HMM profiles for the different kinase catalytic domains and phylogenetic methods [5], yielding a total of 285 protein kinases. Since then, *Ddis* gene models were more extensively curated and HMM kinase profiles further refined and we therefore re-analysed the *Ddis* kinome together with the novel analyses of the more recently sequenced genomes of *Dfas*, *Ppal*, *Dlac*, and *Dpur* [26–28], that represent the four major taxon groups of Dictyostelia. Our analysis was also mainly based on scanning of proteomes with kinase domain specific HMM profiles, combined with

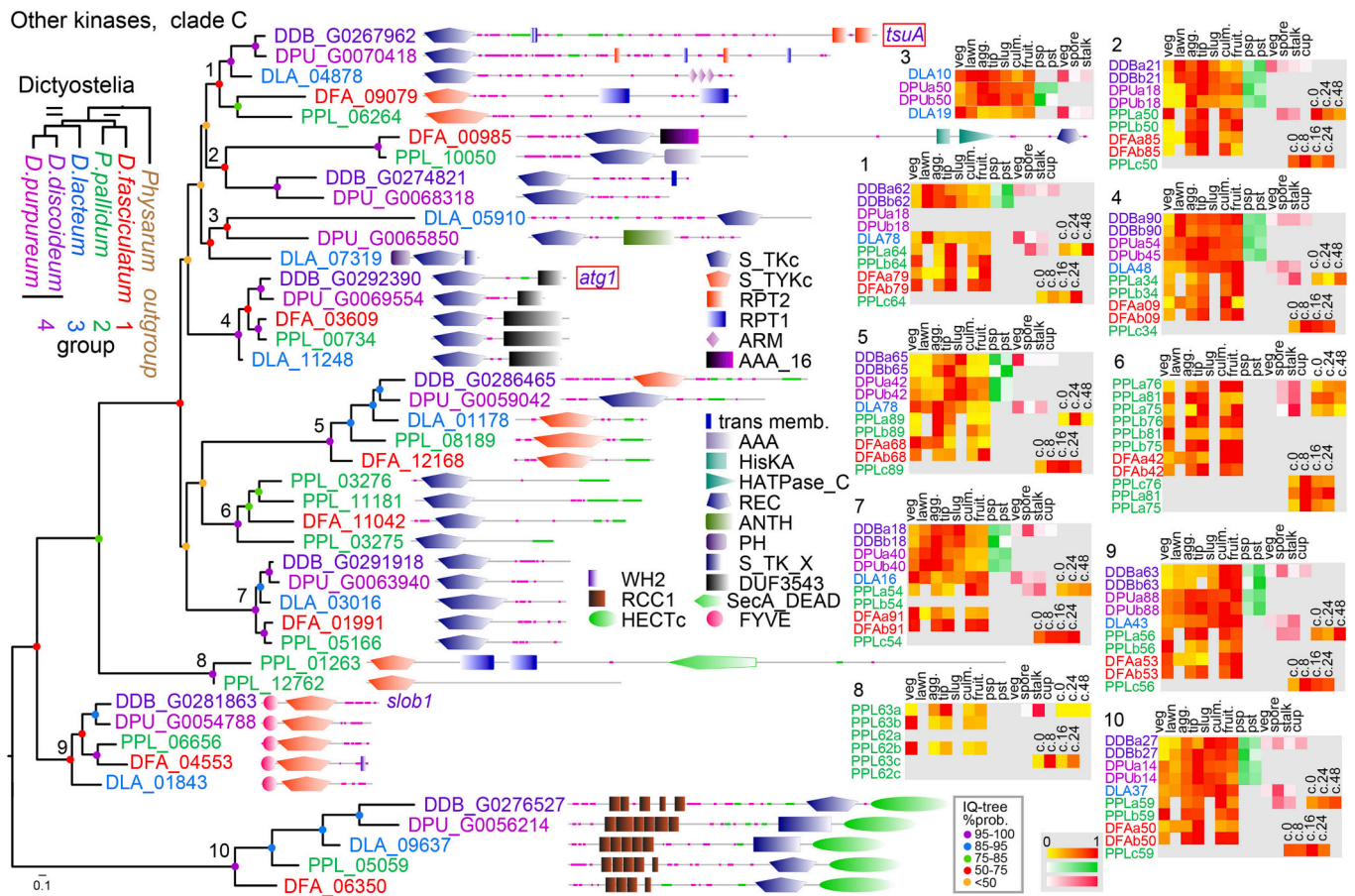


Fig. 2. Example annotated tree of “other” protein kinases, clade C. The groupings of ser/thr/tyr kinases identified in Fig. 1 were subjected individually to phylogenetic inference (Supplemental Figs. S1-S11). After inspection for missing orthologs, the sequences in each grouping were supplemented with hits of BLASTp and tBLASTn queries of genomes, using one of the orthologs as bait. A final tree (here of “other” kinases clade C) was then constructed using maximum-likelihood or bayesian inference-based methods. Gene locus tags are colour-coded to reflect the host species as shown in the species phylogeny (top left) and bootstrap support of the nodes is indicated by coloured dots. The tree was annotated with gene names, which were framed in red for genes with knock-out phenotypes and with the functional domain architecture of the proteins as analysed in SMART [45]. For overlapping domains, we selected the domain with the lowest *E*-value. Clades of orthologous proteins and other groupings were further annotated with heatmaps of relative transcript levels at specific developmental stages or in specific cell types, which were retrieved from published RNA sequencing experiments [28,29,46,47] (yellow-red: 0–1 fraction of maximum value), prespore or prestalk cells (white-green: 0–1 fraction of summed reads), or vegetative, spore, stalk and cup cells (white-red: 0–1 fraction of summed reads). Numbers preceded by c. represent hours of starvation in *Ppal* cells set up for encystation. Note that the phylogeny subdivides the protein kinases in clades of conserved orthologs, with orthology further substantiated by similarity of domain architecture. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sequence selection based on Interpro IDs for protein kinase domains, BLAST search and phylogenetics (see Methods).

The protein kinases can broadly be subdivided into the large group of “typical” ser/thr and tyr kinases with homologous catalytic domains and smaller groupings of “atypical” protein kinases, with different catalytic domains such as the α -kinases, the phosphatidylinositol-3 kinase-related ser/thr protein kinases (PIKKs), the histidine kinases and a few groupings with 4 or less members in Dictyostelia. Their small numbers made the atypical kinases relatively easy to identify and classify. Annotated phylogenies for these kinases are shown in supplemental figs. S12-S15. The initial screens for the typical protein kinases across the five genomes yielded 1242 kinase domains in 1202 proteins, subdivided by HMM scan over the kinase subclasses AGC, CAMK, CK1, CMGC, STE and TKL. A phylogeny inferred from the aligned catalytic domains is shown in Fig. 1, with gene locus tags colour-coded to reflect the HMM profile that yielded the highest probability hit to the sequence. Six monophyletic groupings of kinase domains belonging to the same subtype clearly stand out and in addition there are five additional groupings that contain a mixture of kinase subclasses. These are the so-called “other” kinases, which largely represent low probability hits to specific HMM profiles and/or are detected equally well by multiple profiles. Since the “other” kinases represent a fairly large proportion (24%) of the typical kinases and separate phylogenetically into 5 groupings, we labeled these groupings as Clades A-E. The large size of the phylogeny makes it impossible to annotate and inspect it in a legible format and we therefore prepared separate phylogenies of the sequences contained within the 6 major groups and 5 clades. These phylogenies were inspected for missing orthologs, which were then identified (when present) by BLASTp or tBLASTn query of the relevant proteome or genome and final phylogenies were prepared (Figs. S1-S11). All phylogenies were annotated with the functional domain architectures of the full proteins and heatmaps of the relative expression data of the genes during the developmental cycle of the five species and in specific cell-types. (Fig. 2 and Figs. S1-S15). The supplemental figures are presented with a brief description of each kinase subtype and with phenotypic consequences of gene knock-out and inferred biological roles of individual kinases, when reported.

3.2. Conservation and change in protein kinases across Dictyostelia

Comparisons of the *Ddis* protein kinase repertoire and those of vertebrates and other eukaryotes were reported previously [5,35]. Here, we therefore restricted ourselves to comparisons between representatives of the major dictyostelid taxon groups. Table 1 summarizes the number of kinase domains in all typical and atypical groups between *Ddis*, *Dpur*, *Dlac*, *Ppal* and *Dfas* with the previously enumerated *Ddis* kinases included for comparison. The current and earlier *Ddis* kinase analyses found largely comparable numbers of kinase domains in each category. The earlier analysis detected 12 more “other” kinases, of which four may have been gene prediction errors and the others very derived pseudogenes. *Dpur* and the taxon group 1, 2 and 3 representative species have similar number of kinase domains of each subtype, with *Dlac* having a relatively low number of “other” kinase domains.

The data in Table 1 provide no information on the conservation of individual kinases. Such information as well as information on conserved domain architecture and conserved developmental regulation and cell type specific expression was obtained by compiling the data shown from the annotated phylogenies of figs. S1-S15 into supplemental spreadsheet Table S1_PK compilation.xlsx. The various aspects of gene conservation were colour coded and are summarized in Fig. 3. The compiled data also allows identification of evolutionary trends in species-specific gene gain or loss, or changes in protein functional domains and in the developmental regulation and cell-type specificity of genes. When compared with phenotypic differences between species [49,50], such data may provide clues regarding how molecular change in the protein kinases gave rise to phenotypic innovation. The trends in

Table 1

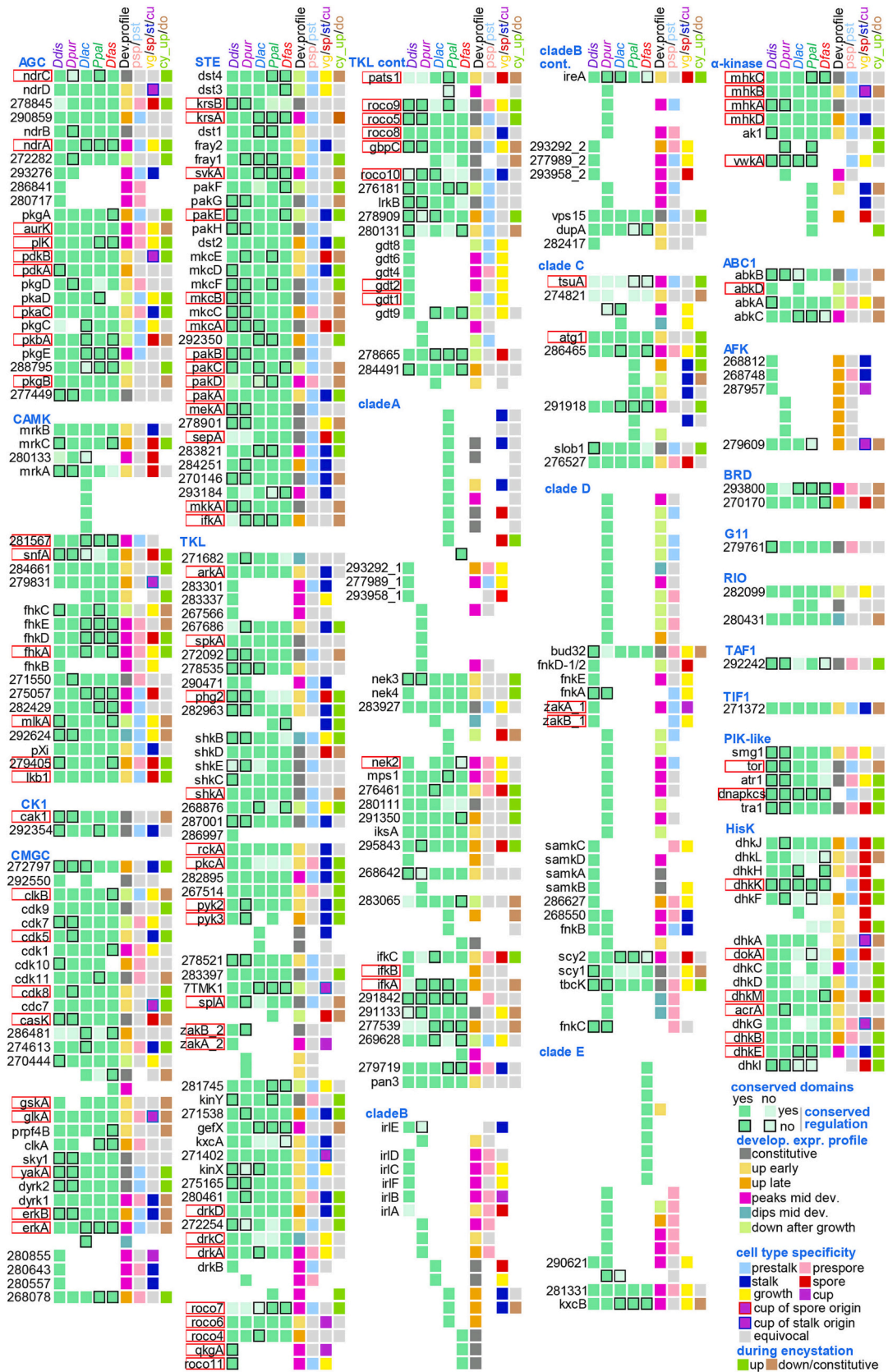
Numbers of protein kinase domains across subtypes and species.

Taxon	subtype	group 4		<i>Dpur</i>	<i>grp.3</i> <i>Dlac</i>	<i>grp.2</i> <i>Ppal</i>	<i>grp.1</i> <i>Dfas</i>	
		<i>Ddis</i>	<i>Ddis</i> Goldberg					
Typical major groups	AGC	24	21	21	21	21	20	
	CAMK	21	21	20	25	20	20	
	CK1	2	2	2	2	2	2	
	CMCG	29	28	24	27	26	26	
	STE	37	44	44	34	34	34	
	TKL	71	68	62	58	61	56	
	cladeA	25		24	21	35	23	
	cladeB	13		14	5	6	6	
	cladeC	7		8	8	12	8	
	cladeD	11		13	4	4	4	
	cladeE	3		9	3	2	12	
	Typical “other”	all “other”	59	71	68	41	59	53
		α -kinase	6		8	6	11	5
		ABC1	4		3	4	3	3
		AFK	4		4	1	1	0
BRD		2		2	2	2	2	
G11		1		1	1	1	1	
RIO		2		2	3	3	2	
PIKK		5		5	5	5	5	
histidine kinase		15		15	14	15	14	
all								
Atypical total	atypical	39	37	281	244	264	243	

Number of domains of the different protein kinase subtypes detected across Dictyostelia. The numbers identified in *Ddis* by Goldberg and coworkers [65] are shown for comparison. Of the larger number detected by Goldberg in the “other” kinase group, four carry Dictybase IDs that no longer exist, while nine are pseudogenes.

gene conservation, cell-type specificity and developmental regulation are shown in Fig. S16A-F for each kinase subtype, while the phylogenetic distribution of conserved features is shown in Fig. S16G-I. The cell-type specific expression of the different kinase subtypes is compared with that of other major families of signal transduction proteins in Fig. 4.

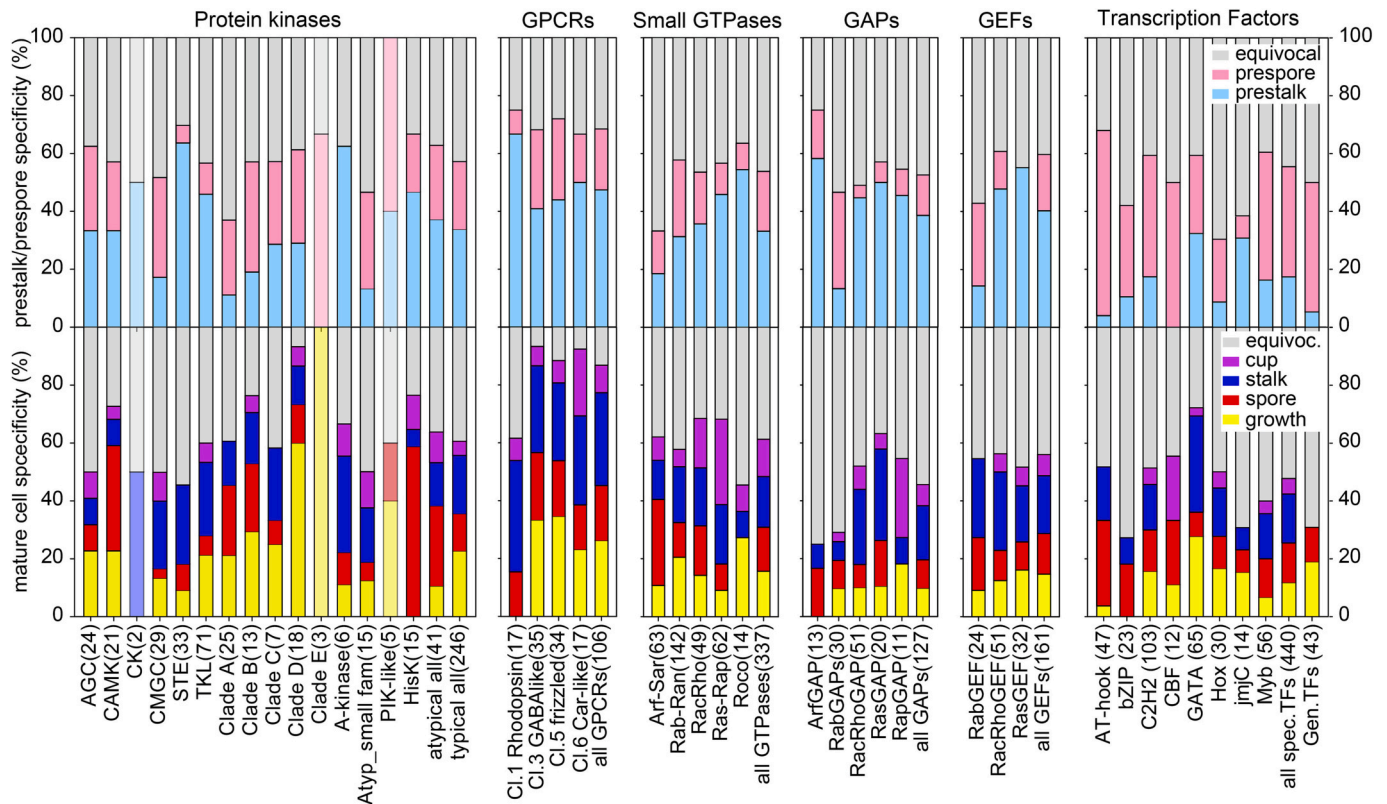
Fig. 3 and Fig. S16A shows that the kinases in the AGC, CAMK, CK1, CMCG, STE and TKL groups are in general (over 70%) conserved across the *Ddis*, *Dpur*, *Dlac*, *Ppal* and *Dfas*. However, the kinases in clades A to E are less conserved (10–40%), which is largely due to extensive species-specific gene amplification. Among the atypical kinases, the PIK-like kinases were fully conserved, while the histidine kinases were missing orthologs in some species. Most protein kinases with known functions in *Ddis* were conserved throughout Dictyostelia, while four (*VwkA*, *DhkI*, *DhkA* and *DhkC*) were missing from either *Dlac* or *Dfas*. *VwkA* is required for contractile vacuole function [67] and lack of this essential gene could be an artifact due to incomplete genome sequence. The *Dhk* histidine kinases are sensors and transducers for developmental signals such as spore differentiation factor 2 (*DhkA*) and ammonia (*DhkC*) [68, 69] and their absence from individual species may mean that such species do not use these signals. Some kinases, such as *IfkB*, *Roco11*, *QkgA* and *ZakA* and *ZakB* are only present in group 4. *IfkB* has an overlapping role with *IfkA* in proper *Ddis* morphogenesis [70] and may improve robustness to development. *Roco11* negatively affects stalk length [71], but since stalk length is highly variable between species, its appearance in group 4 cannot be construed as a cause for phenotypic innovation. *QkgA* acts downstream of the proliferation inhibitors *AprA* and *CfaD* [72], which are conserved throughout Dictyostelia. However, *GrlH*, the *AprA* receptor is also unique to group 4 [73][30] suggesting that *AprA* (and *CfaD*) may have acquired novel roles in group 4. Most interesting is the unique presence of *ZakA* and *ZakB* in group 4. These protein kinases act downstream of GSK3 (glycogen synthase kinase 3) in a pathway where cAMP acting on GSK3 prevents DIF-1 induced trans-differentiation of prespore cells into basal disc cells [74,75]. Since the basal disc is a group 4 novelty [49] and GSK3 activity does not affect



(caption on next page)

Fig. 3. Summary of conservation and change in all analysed protein kinases.

The information contained in all annotated protein kinase subfamily trees (Figs. S1-S15) was collated in supplemental Table S1 (TableS1_PKcompilation.xlsx), colour coded and presented here in summary format. The presence of orthologous protein kinases across the *Ddis*, *Dpur*, *Dlac*, *Ppal* and *Dfas* genomes is indicated by green squares in the first 5 columns, which are shown in pale green or with a black border, respectively, when compared to the majority, the functional domains or the developmental regulation are not conserved. Where the number of non-conserved features is larger than 3, pale green or a border is applied to all squares. The colour coding of the 6th, 7th and 8th square in each row respectively represent the developmental expression profile in the majority of species, the prestalk/prespore specificity, when conserved between *Ddis* and *Dpur* slugs, the growth, spore or stalk specificity, when conserved between species, and the cup cell specificity in *Ddis*. The 9th square represents up- or down regulation in encystation of *Ppal*. Cup cells are only present in group 4 and are bordered red or blue when the orthologs in group 2 or 3 show spore- or stalk-specific expression, respectively. Grey reflects lack of specificity or conflicting data between species or replicate experiments, and white reflects absence of gene or data. The genes are listed by the *Ddis* gene names or 12-digit Dictybase gene identifiers without the DDB_G0 prefix. Genes with known function in *Ddis* are bordered in red. Multiple kinase domains in the same protein are labeled with ₁, ₂ or ₃ in N- to C-terminal order. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Fig. 4.** Cell-type specificity profiles of signal transduction protein families.

The top row shows the percentage of proteins that are specifically expressed in the prestalk or prespore cells of *Ddis* and *Dpur* slugs for all kinase subtypes and for previously investigated large families of signal transduction proteins, such as the G-protein coupled receptors (GPCRs) [30], the small GTPases and their activating (GEF) and inactivating (GAP) proteins [31] and the transcription factors [29]. The bottom row shows the percentage of proteins expressed in mature spore, stalk, and cup cells and vegetative cells in the majority of tested species (*Ddis*, *Dlac* and *Ppal*) for the same families. The name of each subfamily and the number of its non-orthologous members throughout Dictyostelia are shown at the X-axis. For the protein kinases, the figure is based on the data compiled in Supplemental Table S1. For families with <6 members, the bars are shown in washed-out colour.

prespore differentiation in non-group 4 species [76], this highlights co-option of GSK3 in a novel ZakA and ZakB mediated pathway that regulates differentiation of a novel cell-type.

Gene conservation was largely inclusive of the full domain architecture. Since most kinases only consisted of their single kinase domain, this feature lacked distinctiveness. However, consistency in the presence and relative locations of domains did provide strong consolidating evidence of gene orthology (see e.g. Fig. 2). Developmental regulation was often not conserved. While this could affect single species or species scattered across the phylogeny, differences between *Ddis* and *Dpur* (group 4) on one hand and *Dfas*, *Ppal* and *Dlac* (groups 1–3) on the other were more than twice as prevalent as those between the genetically more diverse groupings of *Ddis*, *Dpur* and *Dlac* (Branch II) and *Ppal* and *Dfas* (Branch I)(Fig. S16I). This was also observed in analyses of transcription factor and small GTPase gene families and correlates with

group 4 being phenotypically the most distinctive of all groups [49,50].

With most of the largest families of regulatory proteins now having been analysed [29–31], we considered it of interest to compare the cell-types in which these families and their subfamilies are predominantly expressed. For *Ddis*, such data is available for the prestalk and prespore cells in migrating slugs [46], for the mature stalk, spore and cup cells in the fruiting body and for growing cells [47]. Fig. 4 shows that prespore cells in slugs expressed many more transcription factors than the prestalk cells, except those of the jumonjiC family. However, in stalk and cup cells (which are derived from a prestalk sub-population) more transcription factors were expressed than in spores. Abundant transcription factor expression in prespore cells, but not spores, likely reflects that after aggregation the prespore cells start to prefabricate components and the first layer of the spore wall in numerous Golgi-derived vesicles. Spore maturation mostly involves rapid exocytosis of

these vesicles and further completion of the wall [51]. This means that most gene expression and therefore transcription factor activity for sporulation occurs early in the slug. The stalk only starts to be gradually formed from the onset of fruiting body formation, while cup cells differentiate even later, hence the late requirement for transcription factor activity in stalk and cup differentiation.

Conversely, members of the GPCRs and the Rac/Rho and Ras/Rap families of small GTPases and their GEFs and GAPs were more frequently expressed in prestalk than prespore cells, while later the stalk and cup cells retained their high level of expression (Fig. 4). These families particularly regulate remodelling of the actin cytoskeleton during cell migration and chemotaxis [52,53], indicating that prestalk cells are more specialized in morphogenetic signalling and cell movement. The protein kinases are overall slightly prestalk enriched, with the STE and TKL subtypes and the α -kinases being most prestalk enriched, and the CMGC and “other” kinases showing prespore enrichment. The histidine kinases undergo a dramatic shift from being prestalk enriched in slugs to strongly spore enriched in fruiting bodies.

Likely, their overall mixed and less pronounced affinities for specific cell-types reflect that the protein kinases regulate both gene expression and cytoskeletal remodelling as well as many other cellular processes.

3.3. Cluster analysis of signal transduction proteins

Apart from the comparative analysis of the aforementioned families of transcription factors, GPCRs, GTPases and GAPs and GEFs, we previously also performed an evolutionary comparative analysis of all *Ddis* genes (385 in total), which upon lesion cause developmental defects [28]. While for several of these genes upstream regulators or downstream effectors were also identified, few complete signalling pathways have been resolved. Because proteins that interact in a signalling pathway have to be expressed at the same developmental stage and in the same cell type, similarities in the developmental and cell-type specific transcription profiles of genes can assist in identifying interacting components in a pathway or process. Hierarchical clustering based on correlations between transcription profiles conveniently orders genes with similar expression in a cladogram. We recently explored the utility of hierarchical clustering to identify interactions between GTPases and their potential activators the GEFs (guanine nucleotide exchange factors) and inactivators, the GAPs (GTPase activating proteins) [31]. Within these groups many interactions have already been experimentally established, which allowed us to select the subsets of transcriptome data, the distance metric and the linkage method that returned most of the experimentally established interactions. We here use the same set of transcriptome data for *Ddis*, *Dpur*, *Dlac*, *Ppal* and *Dfas* as in the previous study and could therefore restrict ourselves to already established optimal parameters, which were Pearson correlation as the distance metric and average linkage to order the genes into clusters. The transcription profiles were the same as used for the heatmaps in Fig. 2 and Figs. S1-S15, except that data from individual experiments were used for the cell-type specificity profiles instead of averages. Including transcriptomes from more species in the analysis improves its resolution but can also increase error when in some species the expression of individual genes responded differentially to selection in their specific habitats. Non-group 4 species also do not develop as synchronously and as temporally consistent between experiments as those in group 4, which renders their developmental transcription profiles more noisy. We found in the earlier study that using only the transcriptome data from group 4 or branch II returned more experimentally confirmed interactions than using those from the full phylogeny. However, the greater majority of genes that clustered together in e.g. the group 4 based analysis were also clustered together in analyses of the other subsets of transcriptome data [31].

Fig. 5 shows an overview of the group 4 based cluster analysis annotated with the transcription profiles of the group 4 species. The figure shows that genes with similar developmental and cell-type specific

expression form distinct clusters, but due to the large number of incorporated genes the figure cannot provide further detail. This analysis and the analyses based on transcription profiles of branch II and of the full phylogeny are therefore also archived in Supplemental spreadsheet TableS2_ClusterAnalysis.xlsx. In this spreadsheet, the hierarchical trees are annotated with gene names and locus tags and with the relative gene expression levels upon which the distance calculations were based. They allow inspection of the individual clusters for putative interactors to a gene of interest.

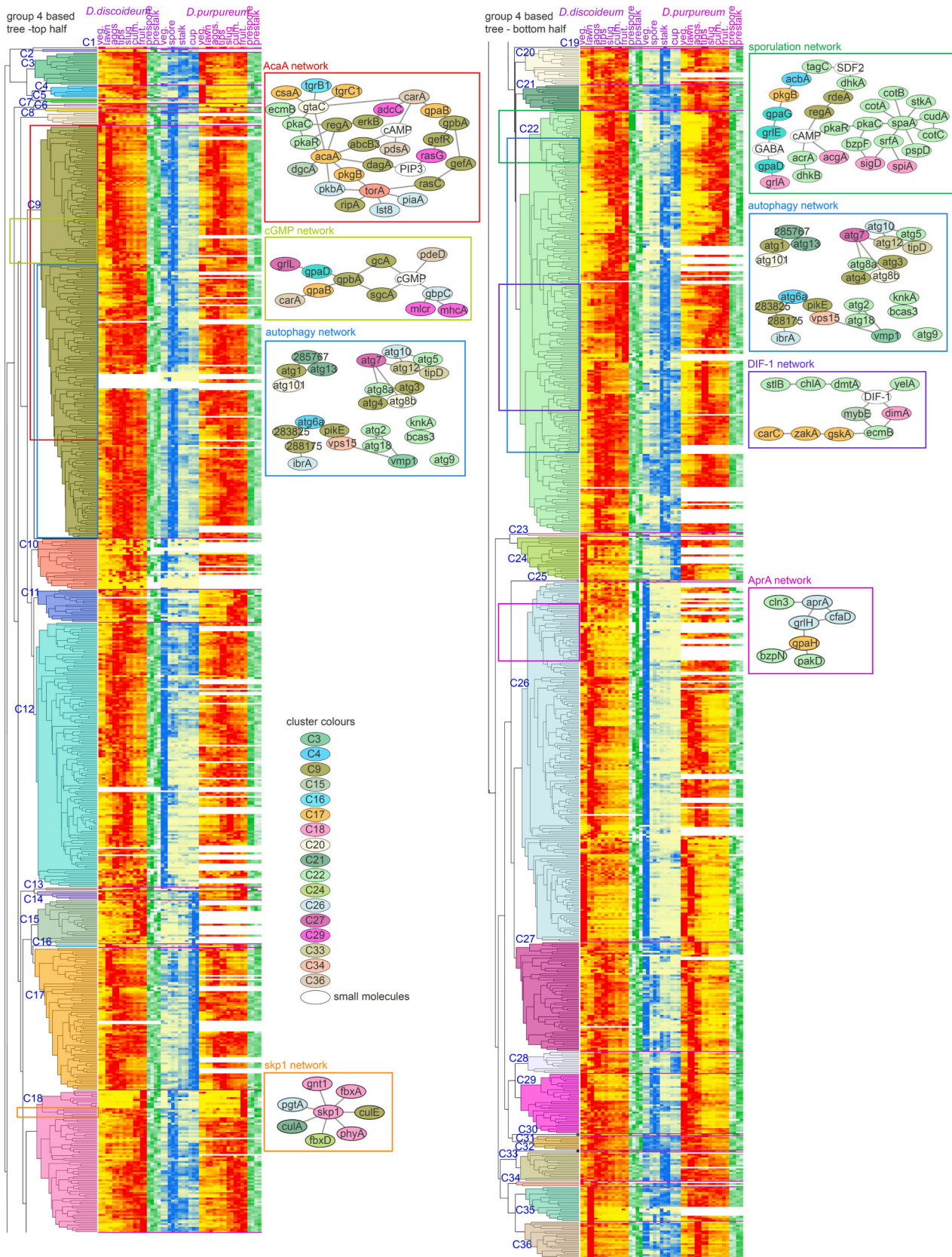
To validate similarity of gene expression as a metric for putative gene interaction, we retrieved literature data on sets of experimentally established interacting proteins. These sets are listed in TableS2_ClusterAnalysis.xlsx, sheet “networks” and visualized as networks with Cytoscape (<https://cytoscape.org/>). The network nodes are coloured similarly as the cluster where the gene resides and the networks are positioned adjacent to the cluster that contain most of its components.

For none of the networks were all components contained in a single cluster, but many subsets of interacting genes were clustered. For the autophagy [54] and AcaA regulatory networks [55,56], genes were dispersed over most of the large clusters, suggesting only loose associations between their transcription profiles. However, the majority of genes in the sporulation network [57,58] resided in a small subclade of cluster 22. Equally, several components of the DIF-1 [59], AprA [60,61] and cGMP networks [62] are contained within subclades of clusters 22, 25 and 9. Three components of the oxygen sensing Skp1 network [63] are members of a co-regulated clade of only five genes in cluster 18.

While interacting proteins naturally need to be expressed at the same time and place, there can be many reasons why the transcription profiles of their cognate genes do not strongly overlap. mRNA or protein stability between members of a network may vary, requiring different patterns of gene expression, or the activity of some members may be strongly affected by post-translational modification. In addition, a constitutively expressed component may interact with several differentially expressed but functionally homologous components and the network itself may undergo stage- or celltype specific functional adaptations. These factors all impact the accuracy of predicting protein interactions from gene co-expression. Nevertheless, our examples also show that profiles of subsets of interactors do cluster together, indicating that gene co-expression can predict protein-protein interaction. However, it is advisable to use co-expression only to support evidence from other approaches such as e.g. provided in the STRING database [64] or from proteomics before using it as the basis for extensive experimentation.

4. Conclusions

- We investigated the presence of protein kinases from all recognized families across genomes that represent the four major taxon groups of Dictyostelia, with two genomes (*Ddis* and *Dpur*) representing group 4.
- Phylogenetic trees of the 19 identified kinase subtypes were inferred and annotated with the functional domain architecture of the proteins and with the developmental- and cell type specific expression profiles of their genes across the five studied taxa.
- For protein kinases that were studied previously in Dictyostelia, information on their likely function and null mutant phenotype was collated and briefly summarized.
- The genomes contained from 243 to 282 different protein kinase domains, with the highest number found in group 4
- Most protein kinases were conserved across all taxon groups, apart from the “other” subtype of typical kinases and the alpha and AFK kinases among the atypical groups, which showed extensive taxon-specific single gene amplifications.
- The developmental and cell-type specific gene expression profiles of the protein kinases were combined with those of other large families of signal transduction proteins and of 385 genes that have essential



(caption on next page)

Fig. 5. Gene expression based hierarchical clustering of signal transduction proteins.

The developmental- and cell type specific expression profiles of protein kinases, GPCRs, small GTPases and their GEFs and GAPs, transcription factors and developmentally essential genes of Dictyostelia of the Group 4 species *Ddis* and *Dpur* were ordered by hierarchical clustering. Pearson correlation was used to estimate distances between the profiles and average linkage to order the clusters. Heatmaps of the expression profiles of all clustered genes are shown. Clusters (C) at a relative branch height of 68% are arbitrarily colour coded to aid inspection of similarity of gene expression in individual clusters. The figure is horizontally split in two to aid presentation. Experimentally inferred networks of interacting proteins are visualized in Cytoscape <https://cytoscape.org/> with the ovals at the nodes (genes) bearing the same colour as the cluster that contains their transcription profile. The positions within the cluster of all network components are contained inside a coloured rectangle. Lists of interacting proteins and the complete hierarchical cluster annotated with locus tags, gene names and transcription data is shown in Supplemental Table S2.ClusterAnalysis.xlsx, which also contains cluster analyses based on transcription profiles of branch II (*Ddis*, *Dpur* and *Dlac*) and across the full phylogeny (groups 1–4).

roles in *Ddis* development. The dataset was subjected to hierarchical clustering to identify sets of co-regulated genes that potentially interact in a regulatory network.

- The cluster analysis brought together known interacting proteins from subsets of experimentally established interaction networks.

Funding

This work was funded by grant 100293/Z/12/Z from the Wellcome Trust, grant 742288 from the European Research Council and grant BB/K000799/1 from the BBSRC. K.K. was also supported by EMBO Long-term fellowship ALTF 295–2015 and by JSPS Overseas Research Fellowship H28–1002. The funding sources had no involvement in the study design, data collection and analysis, manuscript writing and decision to submit the work for publication.

Author contributions

R.S., K.K. G.J.B and C.C. identified and classified ePK sequences across Dictyostelia, K.K. inferred the ePK phylogeny and phylogenies for ePK subtypes and atypical PKs. G.F., Z-H. C., H.L., C.S. and P.S. identified atypical PKs, curated datasets, annotated phylogenies and prepared summary reviews of kinase subsets. P.S. and G.J.B. designed the study and P.S. wrote the manuscript, which was read and corrected where necessary by all co-authors.

Data deposition

All data generated in this work are presented in the main text, supplemental information and the supplemental spreadsheets.

CRedit authorship contribution statement

Koryu Kin: Investigation, Methodology, Formal analysis, Visualization, Writing – original draft. **Zhi-hui Chen:** Investigation, Formal analysis, Visualization, Writing – original draft. **Gillian Forbes:** Investigation, Formal analysis, Visualization, Writing – original draft. **Hajara Lawal:** Investigation, Formal analysis, Visualization, Writing – original draft. **Christina Schilde:** Investigation, Formal analysis, Visualization, Writing - original draft. **Christian Cole:** Methodology, Supervision. **Geoffrey J. Barton:** Conceptualization, Methodology, Supervision. **Pauline Schaap:** Conceptualization, Methodology, Validation, Writing – review & editing, Funding acquisition, Supervision.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Data availability

All data generated in this work is made available in the supplementary materials

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cellsig.2023.110714>.

References

- [1] P. Cohen, D. Cross, P.A. Jänne, Kinase drug discovery 20 years after imatinib: progress and future directions, *Nat. Rev. Drug Discov.* 20 (7) (2021) 551–569.
- [2] S.K. Hanks, T. Hunter, Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification, *FASEB J.* 9 (8) (1995) 576–596.
- [3] G. Manning, D.B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, The protein kinase complement of the human genome, *Science* 298 (5600) (2002) 1912–1934.
- [4] C.A. Bradham, K.R. Foltz, W.S. Beane, M.I. Arnone, F. Rizzo, J.A. Coffman, A. Mushegian, M. Goel, J. Morales, A.M. Genevieve, F. Lapraz, A.J. Robertson, H. Kelkar, M. Loza-Coll, I.K. Townley, M. Raisch, M.M. Roux, T. Lepage, C. Gache, D.R. McClay, G. Manning, The sea urchin kinome: a first look, *Dev. Biol.* 300 (1) (2006) 180–193.
- [5] J.M. Goldberg, G. Manning, A. Liu, P. Fey, K.E. Pilcher, Y.J. Xu, J.L. Smith, The Dictyostelium kinome - analysis of the protein kinases from a simple model organism, *PLoS Genet.* 2 (3) (2006) 291–303.
- [6] 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 (2004) 79.
- [7] D. Miranda-Saavedra, M.J. Stark, J.C. Packer, C.P. Vivares, C. Doerig, G.J. Barton, The complement of protein kinases of the microsporidium *Encephalitozoon cuniculi* in relation to those of *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, *BMC Genomics* 8 (2007) 309.
- [8] M. Zulawski, G. Schulze, R. Braginets, S. Hartmann, W.X. Schulze, The Arabidopsis Kinome: phylogeny and evolutionary insights into functional diversification, *BMC Genomics* 15 (1) (2014) 548.
- [9] J. Middelbeek, K. Clark, H. Venselaar, M.A. Huynen, F.N. van Leeuwen, The alpha-kinase family: an exceptional branch on the protein kinase tree, *Cell. Mol. Life Sci.* 67 (6) (2010) 875–890.
- [10] P.M. Wolanin, P.A. Thomason, J.B. Stock, Histidine protein kinases: key signal transducers outside the animal kingdom, *Genome Biol.* 3 (10) (2002) 3013.
- [11] G.C.M. Smith, S.P. Jackson, Chapter 77 - the PIKK family of protein kinases, in: R. A. Bradshaw, E.A. Dennis (Eds.), *Handbook of Cell Signaling* (Second Edition), Academic Press, San Diego, 2010, pp. 575–580.
- [12] C.J. Pears, J. Brustel, N.D. Lakin, Dictyostelium discoideum as a model to assess genome stability through DNA repair, *Front Cell Dev Biol* 9 (2021), 752175.
- [13] J. Kuhn, Y. Lin, P.N. Devreotes, Using live-cell imaging and synthetic biology to probe directed migration in Dictyostelium, *Front Cell Dev Biol* 9 (2021), 740205.
- [14] R. Graf, M. Grafe, I. Meyer, K. Mitic, V. Pitzten, The Dictyostelium centrosome, *Cells* 10 (10) (2021).
- [15] M. Maniak, Dictyostelium as a model for human lysosomal and trafficking diseases, *Semin. Cell Dev. Biol.* 22 (1) (2011) 114–119.
- [16] R.S.B. Williams, K. Boeckeler, R. Graf, A. Muller-Taubenberg, Z.R. Li, R.R. Isberg, D. Wessels, D.R. Soll, H. Alexander, S. Alexander, Towards a molecular understanding of human diseases using Dictyostelium discoideum, *Trends Mol. Med.* 12 (9) (2006) 415–424.
- [17] C.L. Storey, R.S.B. Williams, P.R. Fisher, S.J. Annesley, Dictyostelium discoideum: a model system for neurological disorders, *Cells* 11 (3) (2022).
- [18] J. Martín-González, J.F. Montero-Bullón, J. Lacal, Dictyostelium discoideum as a non-mammalian biomedical model, *Microb. Biotechnol.* 14 (1) (2021) 111–125.
- [19] R. Escalante, J.J. Vicente, Dictyostelium discoideum: a model system for differentiation and patterning, *Int J Dev Biol* 44 (8) (2000) 819–835.
- [20] K. Kin, P. Schaap, Evolution of multicellular complexity in the Dictyostelid social amoebas, *Genes* 12 (4) (2021) 487.
- [21] P. Kundert, G. Shaulsky, Cellular allorecognition and its roles in Dictyostelium development and social evolution, *Int J Dev Biol* 63 (8–9–10) (2019) 383–393.
- [22] J.M. Medina, P.M. Shreenidhi, T.J. Larsen, D.C. Queller, J.E. Strassmann, Cooperation and conflict in the social amoeba Dictyostelium discoideum, *Int J Dev Biol* 63 (8–9–10) (2019) 371–382.
- [23] P. Schaap, T. Winckler, M. Nelson, E. Alvarez-Curto, B. Elgie, H. Hagiwara, J. Cavender, A. Milano-Curto, D.E. Rozen, T. Dingermann, R. Mutzel, S.L. Baldauf, Molecular phylogeny and evolution of morphology in the social amoebas, *Science* 314 (5799) (2006) 661–663.

- [24] C. Schilde, H.M. Lawal, K. Kin, I. Shibano-Hayakawa, K. Inouye, P. Schaap, A well supported multi gene phylogeny of 52 dictyostelia, *Mol. Phylogenet. Evol.* 134 (2019) 66–73.
- [25] L. Eichinger, J.A. Pachebat, G. Glockner, M.A. Rajandream, R. Sugchang, M. Berriman, J. Song, R. Olsen, K. Szafarski, Q. Xu, B. Tunggal, S. Kummerfeld, M. Madera, B.A. Konfortov, F. Rivero, A.T. Bankier, R. Lehmann, N. Hamlin, R. Davies, P. Gaudet, P. Fey, K. Pilcher, G. Chen, D. Saunders, E. Sodergren, P. Davis, A. Kerhornou, X. Nie, N. Hall, C. Anjard, L. Hemphill, N. Bason, P. Farbrother, B. Desany, E. Just, T. Morio, R. Rost, C. Churcher, J. Cooper, S. Haydock, N. van Driessche, A. Cronin, I. Goodhead, D. Muzny, T. Mourier, A. Pain, M. Lu, D. Harper, R. Lindsay, H. Hauser, K. James, M. Quiles, M. Madan Babu, T. Saito, C. Buchrieser, A. Wardroper, M. Felder, M. Thangavelu, D. Johnson, A. Knights, H. Loulseged, K. Mungall, K. Oliver, C. Price, M.A. Quail, H. Urushihara, J. Hernandez, E. Rabinowitz, D. Steffen, M. Sanders, J. Ma, Y. Kohara, S. Sharp, M. Simmonds, S. Spiegler, A. Tivey, S. Sugano, B. White, D. Walker, J. Woodward, T. Winckler, Y. Tanaka, G. Shaulsky, M. Schleicher, G. Weinstock, A. Rosenthal, E.C. Cox, R.L. Chisholm, R. Gibbs, W.F. Loomis, M. Platzer, R.R. Kay, J. Williams, P.H. Dear, A.A. Noegel, B. Barrell, A. Kuspa, The genome of the social amoeba *Dictyostelium discoideum*, *Nature* 435 (7038) (2005) 43–57.
- [26] A. Heide, H. Lawal, M. Felder, C. Schilde, N. Helps, B. Tunggal, F. Rivero, U. John, M. Schleicher, L. Eichinger, M. Platzer, A. Noegel, K. Schaap, G. Glockner, Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication, *Genome Res.* 21 (2011) 1882–1891.
- [27] R. Sugchang, A. Kuo, X. Tian, W. Salerno, A. Parikh, C.L. Feasley, E. Dalin, H. Tu, E. Huang, K. Barry, E. Lindquist, H. Shapiro, D. Bruce, J. Schmutz, A. Salamov, P. Fey, P. Gaudet, C. Anjard, M.M. Babu, S. Basu, Y. Bushmanova, H. van der Wel, M. Katoh-Kurasawa, C. Dinh, P.M. Coutinho, T. Saito, M. Elias, P. Schaap, R.R. Kay, B. Henrissat, L. Eichinger, F. Rivero, N.H. Putnam, C.M. West, W.F. Loomis, R. L. Chisholm, G. Shaulsky, J.E. Strassmann, D.C. Queller, A. Kuspa, I.V. Grigoriev, Comparative genomics of the social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*, *Genome Biol.* 12 (2) (2011) R20.
- [28] G. Gloeckner, H.M. Lawal, M. Felder, R. Singh, G. Singer, C.J. Weijer, P. Schaap, The multicellularity genes of dictyostelid social amoebas, *Nat. Commun.* 7 (2016) 12085.
- [29] G. Forbes, Z.H. Chen, K. Kin, H.M. Lawal, C. Schilde, Y. Yamada, P. Schaap, Phylogeny-wide conservation and change in developmental expression, cell-type specificity and functional domains of the transcriptional regulators of social amoebas, *BMC Genomics* 20 (1) (2019) 890.
- [30] G. Hall, S. Kelly, P. Schaap, C. Schilde, Phylogeny-wide analysis of G-protein coupled receptors in social amoebas and implications for the evolution of multicellularity [version 2; peer review: 2 approved], *Open Research Europe* 2 (134) (2023).
- [31] G. Forbes, C. Schilde, H. Lawal, K. Kin, Q. Du, Z.H. Chen, F. Rivero, P. Schaap, Interactome and evolutionary conservation of Dictyostelid small GTPases and their direct regulators, *Small GTPases* (2021) 1–16.
- [32] N. Bhardwaj, H. Lu, Correlation between gene expression profiles and protein-protein interactions within and across genomes, *Bioinformatics* 21 (11) (2005) 2730–2738.
- [33] S. De Bodt, S. Proost, K. Vandepoele, P. Rouzé, Y. Van de Peer, Predicting protein-protein interactions in Arabidopsis thaliana through integration of orthology, gene ontology and co-expression, *BMC Genomics* 10 (2009) 288.
- [34] M.B. Eisen, P.T. Spellman, P.O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci. U. S. A.* 95 (25) (1998) 14863–14868.
- [35] D. Miranda-Saavedra, G.J. Barton, Classification and functional annotation of eukaryotic protein kinases, *Proteins* 68 (4) (2007) 893–914.
- [36] P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A.F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, S. Hunter, InterProScan 5: genome-scale protein function classification, *Bioinformatics* 30 (9) (2014) 1236–1240.
- [37] J. Mistry, R.D. Finn, S.R. Eddy, A. Bateman, M. Punta, Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions, *Nucleic Acids Res.* 41 (12) (2013), e121.
- [38] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Mol. Biol. Evol.* 30 (4) (2013) 772–780.
- [39] S. Capella-Gutiérrez, J.M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, *Bioinformatics* 25 (15) (2009) 1972–1973.
- [40] L.T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, *Mol. Biol. Evol.* 32 (1) (2015) 268–274.
- [41] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermini, ModelFinder: fast model selection for accurate phylogenetic estimates, *Nat. Methods* 14 (6) (2017) 587–589.
- [42] F. Sievers, D.G. Higgins, Clustal omega, accurate alignment of very large numbers of sequences, *Methods Mol. Biol.* 1079 (2014) 105–116.
- [43] F. Ronquist, J.P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models, *Bioinformatics* 19 (12) (2003) 1572–1574.
- [44] A. Bateman, E. Birney, R. Durbin, S.R. Eddy, K.L. Howe, E.L.L. Sonnhammer, The Pfam protein families database, *Nucleic Acids Res.* 28 (1) (2000) 263–266.
- [45] J. Schultz, F. Milpetz, P. Bork, C.P. Ponting, SMART, a simple modular architecture research tool: identification of signaling domains, *Proc. Natl. Acad. Sci. U. S. A.* 95 (11) (1998) 5857–5864.
- [46] A. Parikh, E.R. Miranda, M. Katoh-Kurasawa, D. Fuller, G. Rot, L. Zagar, T. Curk, R. Sugchang, R. Chen, B. Zupan, W.F. Loomis, A. Kuspa, G. Shaulsky, Conserved developmental transcriptomes in evolutionarily divergent species, *Genome Biol.* 11 (3) (2010) R35.
- [47] K. Kin, G. Forbes, A. Cassidy, P. Schaap, Cell-type specific RNA-Seq reveals novel roles and regulatory programs for terminally differentiated Dictyostelium cells, *BMC Genomics* 19 (1) (2018) 764.
- [48] J. Demsar, T. Curk, A. Erjavec, C. Gorup, T. Hovevar, M. Milutinovic, M. Mozina, M. Polajnar, M. Toplak, A. Staric, M. Stajdohar, L. Umek, L. Zagar, J. Zbontar, M. Zitnik, B. Zupan, Orange: data mining toolbox in Python, *J. Mach. Learn. Res.* 14 (2013) 2349–2353.
- [49] M. Romeralo, A. Skiba, A. Gonzalez-Voyer, C. Schilde, H. Lawal, S. Kedziora, J. C. Cavender, G. Glockner, H. Urushihara, P. Schaap, Analysis of phenotypic evolution in Dictyostelia highlights developmental plasticity as a likely consequence of colonial multicellularity, *Proc. Biol. Sci.* 280 (1764) (2013) 20130976.
- [50] C. Schilde, A. Skiba, P. Schaap, Evolutionary reconstruction of pattern formation in 98 Dictyostelium species reveals that cell-type specialization by lateral inhibition is a derived trait, *EvoDevo* 5 (1) (2014) 34.
- [51] C.M. West, Comparative analysis of spore coat formation, structure, and function in Dictyostelium, *Int. Rev. Cytol.* 222 (2003) 237–293.
- [52] F. Rivero, H. Xiong, Rho Signaling in Dictyostelium discoideum, *Int. Rev. Cell Mol. Biol.* 322 (2016) 61–181.
- [53] Y. Liu, J. Lecal, R.A. Firtel, A. Kortholt, Connecting G protein signaling to chemoattractant-mediated cell polarity and cytoskeletal reorganization, *Small GTPases* 9 (4) (2018) 360–364.
- [54] S. Fischer, L. Eichinger, Dictyostelium discoideum and autophagy - a perfect pair, *Int J Dev Biol* 63 (8–9–10) (2019) 485–495.
- [55] B. Pergolizzi, S. Bozzaro, E. Bracco, G-protein dependent signal transduction and ubiquitination in Dictyostelium, *Int. J. Mol. Sci.* 18 (10) (2017).
- [56] W.F. Loomis, Genetic control of morphogenesis in Dictyostelium, *Dev. Biol.* 402 (2) (2015) 146–161.
- [57] C. Anjard, W.F. Loomis, GABA induces terminal differentiation of Dictyostelium through a GABA(B) receptor, *Development* 133 (11) (2006) 2253–2261.
- [58] Y. Yamada, A. Cassidy, P. Schaap, The transcription factor spores absent a is a PKA dependent inducer of Dictyostelium sporulation, *Sci. Rep.* 8 (1) (2018) 6643.
- [59] T. Araki, T. Saito, Small molecules and cell differentiation in Dictyostelium discoideum, *Int J Dev Biol* 63 (8–9–10) (2019) 429–438.
- [60] Y. Tang, Y. Wu, S.E. Herlihy, F.J. Brito-Aleman, J.H. Ting, C. Janetopoulos, R. H. Gomer, An autocrine proliferation regulator regulates Dictyostelium discoideum proliferation and Chemorepulsion using the G protein-coupled receptor GrIH, *mBio* 9 (1) (2018).
- [61] J.E. Phillips, E. Huang, G. Shaulsky, R.H. Gomer, The putative bZIP transcription factor BzpN slows proliferation and functions in the regulation of cell density by autocrine signals in Dictyostelium, *PLoS One* 6 (7) (2011), e21765.
- [62] J.M. Goldberg, E.S. Wolpin, L. Bosgraaf, B.K. Clarkson, P.J.M. Van Haastert, J. L. Smith, Myosin light chain kinase a is activated by cGMP-dependent and cGMP-independent pathways, *FEBS Lett.* 580 (8) (2006) 2059–2064.
- [63] A.W. Boland, E. Gas-Pascual, B.L. Nottingham, H. van der Wel, N.G. Daniel, M. O. Sheikh, C.M. Schafer, C.M. West, Oxygen-dependent regulation of E3(SCF) ubiquitin ligases and a Skp1-associated JmjD6 homolog in development of the social amoeba Dictyostelium, *J. Biol. Chem.* 298 (9) (2022), 102305.
- [64] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, M. Kuhn, P. Bork, L.J. Jensen, C. von Mering, STRING v10: protein-protein interaction networks, integrated over the tree of life, *Nucleic Acids Res.* 43 (Database issue) (2015) D447–D452.
- [65] J.M. Goldberg, L. Bosgraaf, P.J.M. Van Haastert, J.L. Smith, Identification of four candidate cGMP targets in Dictyostelium, *Proc. Natl. Acad. Sci. U. S. A.* 99 (10) (2002) 6749–6754.
- [66] V. Betapudi, T.T. Egelhoff, Roles of an Unconventional Protein Kinase and Myosin II in Amoeba Osmotic Shock Responses, *Traffic* (2009).
- [67] N. Wang, F. Soderbom, C. Anjard, G. Shaulsky, W.F. Loomis, SDF-2 induction of terminal differentiation in *Dictyostelium discoideum* is mediated by the membrane-spanning sensor kinase DhkA, *Mol. Cell. Biol.* 19 (7) (1999) 4750–4756.
- [68] C.K. Singleton, M.J. Zinda, B. Mykytko, P. Yang, The histidine kinase dhkC regulates the choice between migrating slugs and terminal differentiation in *Dictyostelium discoideum*, *Dev. Biol.* 203 (1998) 345–357.
- [69] M. Rai, Y.H. Xiong, C.K. Singleton, Disruption of the ifkA and ifkB genes results in altered cell adhesion, morphological defects and a propensity to form pre-stalk O cells during development of Dictyostelium, *Differentiation* 74 (9–10) (2006) 583–595.
- [70] W.N. van Egmond, P.J. van Haastert, Characterization of the Roco protein family in Dictyostelium discoideum, *Eukaryot Cell* 9(5) (2010) 751–761.
- [71] J.E. Phillips, R.H. Gomer, The Roco kinase QkgA is necessary for proliferation inhibition by autocrine signals in Dictyostelium discoideum, *Eukaryot Cell* 9 (10) (2010) 1557–1565.
- [72] S.A. Kirolos, R.H. Gomer, A chemorepellent inhibits local Ras activation to inhibit pseudopod formation to bias cell movement away from the chemorepellent, *Mol Biol Cell* 33(1) (2022) ar9, 1–13.
- [73] A.J. Harwood, Use of the Dictyostelium stalk cell assay to monitor GSK-3 regulation, *Meth Mol Biol* 469 (2008) 39–44.

- [75] L. Kim, J. Brzostowski, A. Majithia, N.S. Lee, V. McMains, A.R. Kimmel, Combinatorial cell-specific regulation of GSK3 directs cell differentiation and polarity in *Dictyostelium*, *Development* 138 (3) (2011) 421–430.
- [76] Y. Kawabe, T. Morio, Y. Tanaka, P. Schaap, Glycogen synthase kinase 3 promotes multicellular development over unicellular encystation in encysting *Dictyostelia*, *Evodevo* 9 (2018) 12.
- [77] 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(W1) (2016) W242–245.