The Physarum polycephalum Genome Reveals Extensive Use of Prokaryotic Two-component and Metazoan-type Tyrosine Kinase Signaling

Pauline Schaap, Israel Barrantes, Pat Minx, Narie Sasaki, Roger W. Anderson, Marianne Bénard, Kyle K. Biggar, Nicolas E. Buchler, Ralf Bundschuh, Xiao Chen, et al.

Genome Biology and Evolution Advance Access published November 27, 2015 doi:10.1093/gbe/evv237

Main Manuscript

The *Physarum polycephalum* Genome Reveals Extensive Use of Prokaryotic Two-component and Metazoan-type Tyrosine Kinase Signaling

Pauline Schaap[1], Israel Barrantes[2], Pat Minx[3], Narie Sasaki[4], Roger W. Anderson[5], Marianne Bénard[6], Kyle K. Biggar[7], Nicolas E. Buchler[8], Ralf Bundschuh[9], Xiao Chen[10], Catrina Fronick[3], Lucinda Fulton[3], Georg Golderer[11], Niels Jahn[12], Volker Knoop[13], Laura F. Landweber[10], Chrystelle Maric[14], Dennis Miller[15], Angelika A. Noegel[16], Rob Peace[17], Gérard Pierron[14], Taeko Sasaki[4], Mareike Schallenberg-Rüdinger[13], Michael Schleicher[18], Reema Singh[1], Thomas Spaller[19], Kenneth B. Storey[17], Takamasa Suzuki[20], Chad Tomlinson[3], John J. Tyson[21], Wesley C. Warren[3], Ernst R. Werner[11], Gabriele Werner-Felmayer[11], Richard K. Wilson[3], Thomas Winckler[19], Jonatha M. Gott[22], Gernot Glöckner[16,23]*, Wolfgang Marwan[2]*

*Senior authors, listed in alphabetical order

Author details

[1]College of Life Sciences, University of Dundee, DD15EH, UK. [2]Magdeburg Centre for Systems Biology and Institute for Biology, University of Magdeburg, Pfälzerstrasse 5, 39106 Magdeburg, Germany. [3]The Genome Institute, Washington University School of Medicine, St Louis, MO 63108, USA. [4]Department of Biological Sciences, Graduate School of Science, Nagoya University, Furocho, Chikusaku, Nagoya, 464-8602 Aichi, Japan. [5]Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, UK. [6]UPMC Univ Paris 06, Institut de Biologie Paris-Seine (IBPS), CNRS UMR-7622, F-75005, Paris, France. [7]Biochemistry Department, Schulich School of Medicine and Dentistry, Western University, London, ON, N6A 5C1, Canada. [8]Center for Genomic and Computational Biology, Department of Biology, and Department of Physics, Duke University, Durham, NC 27710, USA. [9]Department of Physics, Department of Chemistry & Biochemistry, Division of Hematology, Department of Internal Medicine, Center for RNA Biology, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210-1117, USA. [10]Department of Ecology & Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA. [11]Biological Chemistry, Biocenter, Innsbruck Medical University, 6020 Innsbruck, Austria. [12]Genome Analysis, Leibniz Institute on Aging - Fritz Lipmann Institute (FLI), Beutenbergstraße 11, 07745 Jena, Germany. [13]IZMB - Institut für Zelluläre und Molekulare Botanik, Universität Bonn, Kirschallee 1, 53115 Bonn, Germany. [14]Institut Jacques Monod, CNRS UMR7592,

Université Paris Diderot Paris7, 75013 Paris, France.[15]The University of Texas at Dallas, Biological Sciences, FO 31, 800 West Campbell Road, Richardson, TX 75080, USA. [16] Institute for Biochemistry I, Medical Faculty, University of Cologne, Joseph-Stelzmann-Straße 52, 50931 Cologne, Germany. [17]Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada. [18]Institute for Anatomy III / Cell Biology, BioMedCenter, Ludwig-Maximilians-Universität, Grosshaderner Str. 9, 82152 Planegg-Martinsried, Germany. [19]Institut für Pharmazie, Friedrich-Schiller-Universität Jena, Semmelweisstraße 10, 07743 Jena, Germany. [20]Department of Biological Sciences, Graduate School of Science and JST ERATO Higashiyama Live-holonics Project, Nagoya University, Furocho, Chikusaku, Nagoya, 464-8602 Aichi, Japan. [21]Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA. [22]Center for RNA Molecular Biology, Case Western Reserve University, School of Medicine, 10900 Euclid Ave, Cleveland, OH 44106-4960, USA. [23]Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, 12587 Berlin.

Abstract

Physarum polycephalum is a well-studied microbial eukaryote with unique experimental attributes relative to the other experimental model organisms. It has a sophisticated life cycle with several distinct stages including amoebal, flagellated, and plasmodial cells. It is unusual in switching between open and closed mitosis according to specific life cycle stages. Here we present the analysis of the genome of this enigmatic and important model organism and compare it with closely related species. The genome is littered with simple and complex repeats and the coding regions are frequently interrupted by introns with a mean size of 100 bases. Complemented with extensive transcriptome data, we define ca. 31,000 gene loci, providing unexpected insights into early eukaryote evolution. We describe extensive use of histidine kinase based two-component systems and tyrosine kinase signaling, the presence of bacterial and plant type photoreceptors (phytochromes, cryptochrome, and phototropin) and of plant-type pentatricopeptide repeat proteins, as well as metabolic pathways, and a cell cycle control system typically found in more complex eukaryotes. Our analysis characterizes *P. polycephalum* as a prototypical eukaryote with features attributed to the last common ancestor of Amorphea, i.e. the Amoebozoa and Opisthokonts. Specifically, the presence of tyrosine kinases in *Acanthamoeba* and *Physarum* as representatives of two distantly related subdivisions of Amoebozoa argues against the later emergence of tyrosine kinase signaling in the opisthokont lineage and also against the acquisition by horizontal gene transfer.

Key words: Amoebozoa, tyrosine kinase receptor, two-component system, signaling, phytochrome

Introduction

P. polycephalum belongs to the Amoebozoa, the sister group to the Opisthokonts (*i.e.* fungi and animals) (Cavalier-Smith 2003) which both together form the supergroup Amorphea (Adl, et al. 2012). In the course of its complex life cycle (Fig. 1A), *P. polycephalum* is able to differentiate into various specialized cell types depending on environmental conditions (Burland, et al. 1993). Macroscopic multinucleate plasmodial cells contain a naturally synchronous replicating and differentiating population of nuclei and can grow to tens or even hundreds of cm in size. Individual pieces cut off from a single plasmodium maintain synchrony, providing unique experimental options for single cell biology. The occurrence of open and closed mitosis in amoebae and plasmodia, respectively, is another well-known aspect of its rich cell biology (Aldrich and Daniel 1982; Dove, et al. 1986; Sauer 1982). Accordingly, *P. polycephalum* has served as a classical model organism in cell and developmental biology since the early 1960s and has been used extensively to study cell cycle regulation, cell differentiation (amoeba-flagellate transition, spherulation, sporulation), cell fusion, DNA replication, developmental gene expression, histone modification, sensing and response (e.g. chemotaxis) (for reviews see (Aldrich and Daniel 1982; Burland, et al. 1993; Dove, et al. 1986; Sauer 1982)). More recently, *P. polycephalum* plasmodia have been used for studies ranging from cell biology and biophysics to soft matter computing for path finding, the biosensory control of robots, or the generation of music (Braund and Miranda 2014;

Tero, et al. 2010; Tsuda, et al. 2007). Plasmodial cells also contain thousands of mitochondria, providing the means to study the highly unusual form of insertional RNA editing used to generate functional mitochondrial transcripts.

From the evolutionary biology point of view, *P. polycephalum* is interesting because of its placement into the Amoebozoa, where the genomes of several distantly related representatives (*A. castellanii*, *E. histolytica*, *etc*.) are already known (Fig. 1B). The model organism most closely related to *P. polycephalum* for which a genome sequence has previously been determined is *Dictyostelium discoideum* (Eichinger, et al. 2005). The two model species differ in several important aspects. *D. discoideum* is a cellular slime mold where differentiation occurs in developmentally interacting cells while in the acellular slime mold *P. polycephalum,* a single, multinucleate plasmodial cell differentiates in a manner reminiscent of *Drosophila* embryos or other developing syncytia. A comparison of *P. polycephalum* and *D. discoideum* may help to elucidate the different evolutionary trajectories the members of the two branches, cellular and acellular slime molds, have taken in establishing sophisticated life cycles that are so different.

The established lab strains of *D. discoideum* do not form viable progeny after mating, which hinders genetic experiments, whereas mating of strains and segregation of meiotic progeny in *P. polycephalum* is under ready experimental control. The classical genetics are well-established in *P. polycephalum* and the possibility to experimentally switch the life cycle between haploid and diploid stages (Fig. 1A) is very useful for the generation and analysis of mutants (Dee 1987, 1982). Amoebal strains derived from different natural plasmodial isolates, such as Wis-1 and Wis-2, (see Appendix A of Supplementary Methods, Results and Discussion) allow forward genetic approaches by mapping genes identified in phenotypic screens.

Mendelian analysis based on the co-segregation of phenotype and single nucleotide polymorphisms (SNPs) (Schedl and Dove 1982) can be performed by sequencing pools of cDNAs of phenotype-carrying segregants (Barrantes and Marwan, unpublished results). These high throughput methods combined with Mendelian genetics mitigate current difficulties of reproducibly making transgenic lines. Methods for knocking down gene expression have recently been developed for *P. polycephalum* (Haindl and Holler 2005; Itoh, et al. 2011), opening up additional experimental avenues. Here, we present an overview of the whole genome of *P. polycephalum*, focussing on some of its peculiar features such as its surprisingly extended signaling system.

Materials and Methods

Assembly

Combined sequence reads from different libraries (including 3 kb and 8 kb mate pairs) were assembled using the Newbler software version 2.6 (Margulies et al 2005), and the Newbler contigs were then assembled into scaffolds using Bambus version 2.3 (Pop et al 2004), and a custom transcriptome-guided scaffolding algorithm. The 10.0 draft assembly is comprised of 69,687 scaffolds with an N50 scaffold length of 54,474kb and an N50 contig length of 2.2kb. The assembled coverage is 54.6X, and the assembly spans over 220Mb. We removed from the assembly all contaminating sequences, trimmed vectors (X), and ambiguous bases (N). Additionally, shorter contigs (≤200bp) were removed prior to public release. The draft genome sequence of *P. polycephalum* was deposited in the DDBJ/EMBL/GenBank database (Accession Number ATCM00000000.3). The genome data can also be downloaded from http://www.physarum-blast.ovgu.de/.

Generation of a reference transcriptome

All transcript data were assembled according to their sample origin since previous analyses showed that genes have a wealth of introns and thus alternative splicing could play a role (Glöckner, et al. 2008). This initial analysis yielded 733,443 potential transcripts.

After clustering with usearch (Edgar 2010) we obtained 31,770 clusters each represented by the largest sequence in a cluster. The mean length of transcripts in the thus constructed reference transcriptome is 1264 bases. These transcripts were mapped onto the genome using blat (Kent 2002). Only 485 transcripts could not be mapped to the genome sequence, of which only 90 have a blast score of 250 or higher in a search against the reference sequence database (NCBI July 2013). Fifteen of these 90 appear to be of *P. polycephalum* mitochondrial origin and a further 13 are certainly contaminants from the sequencing process since they have identity values higher than 85 % to some bacteria and mouse sequences. Thus, only a minor fraction of transcripts is not represented by the assembled genome. Manual inspection of some transcript/genome alignments revealed that transcripts frequently span gaps in the scaffolds which account for missing small exons. Inability to assemble genome junctions at these positions is likely caused by the low complexity of intron sequences and the small size of the missing exons. Since contiguity of a sequence is required for accurate gene prediction, we relied heavily on the transcript data for annotation and analysis. The Transcriptome Reference Assembly has been deposited at DDBJ/EMBL/GenBank under the accession GDRG00000000. The version described in this paper is the first version, GDRG01000000. All transcriptome sequence data and files containing the gene loci-transcript conversion, the transcript map, and the automatic annotation of the reference transcriptome can be downloaded from http://www.physarum-blast.ovgu.de/.

Gene prediction and construction of gene loci

For an *ab initio* gene prediction we used Augustus (Stanke, et al. 2004). The transcriptome data were used to train this program according to the manual and to provide splice information for the algorithm. The program itself was run with default values. This way we predicted 47779 protein coding genes. We noticed, however, that the fragmented nature of the genome sequence and missing contiguity within scaffolds led to an overprediction. To get a better estimate of the number of protein coding genes in the *P. polycephalum* genome we tried to combine gene prediction and reference transcriptome for the definition of gene loci. To this end we mapped the reference transcriptome data to the genome. We then fused predicted genes and reference transcripts to gene loci if predicted genes were covered by the same reference transcript and/or if predicted genes were present in the near vicinity (<50 bases apart) and at least one of them was covered or partially covered by a transcript sequence. Predicted genes without transcript coverage and no transcript in the near vicinity (< 50 bp apart) were counted as independent gene loci. In this way we defined 34438 gene loci, of which 17280 are not supported by transcript data (Table 1).

Small RNAs

For the definition of small RNAs we used RNAseq transcript data from a previous experiment (Bundschuh, et al. 2011). Since this specific experiment was performed on mitochondrial preparations, we only expect the more abundant nuclear small RNAs to be observed. However, the absence of poly-A selection and the use of a small RNA kit in library preparation promised a view orthogonal to traditional RNA-Seq approaches.

For detailed information on strains used, preparation of nucleic acids, and bioinformatic methods see Supplementary File 1.

Results and Discussion

Genome properties and reference transcriptome

The *P. polycephalum* genome proved to be highly recalcitrant for assembly owing to abundant and frequently long (mononucleotide or pyrimidine-rich) repeat stretches present in intergenic regions or within its numerous and highly variable intron sequences. Construction of gene models therefore heavily relied on accompanying transcriptome data (see Materials and Methods).

The genetic material used for sequencing the *P. polycephalum* genome was derived from axenic, haploid amoebal cells (strain LU352). The assembly of 454 and paired end Illumina reads with an estimated coverage of 50x (calculating with a genome size of 250 Mb) resulted in scaffolds covering 182 Mb. However, assembling the genome has been challenging due to extremely long stretches of di-, tri-, and tetranucleotide repeats and large homopolymeric tracts, which likely lead to polymerase slippage and premature polymerase termination *in vitro.* The unresolvable sequence patches cause innumerable gaps in the final sequence, resulting in a mean scaffold size of only slightly larger than 2 kb. PacBio sequencing yielded only minimal improvement. We therefore tested whether our assembly captures most of the coding information of the genome by investigating the primary metabolism capacity encoded by the assembly. We found that all expected pathways were entirely present and therefore conclude that the current assembly is complete enough for further analysis.

Gene content and introns

To further evaluate gene content and completeness, we generated transcriptome data sets from multiple life cycle stages and constructed a reference transcriptome (see Materials and Methods).

We next trained the gene prediction program augustus (Stanke, et al. 2004) with sequences derived from mapped transcripts. Gene predictions using the splicing information from all transcripts initially yielded more than 47,000 gene models (Table 1). However, since in the current assembly some shorter introns and exons are missing in scaffolds due to simple repeat sequences at both ends, we used the reference transcripts together with the gene predictions to define gene loci. Transcripts and predicted genes overlapping each other or being in close vicinity on the assembly (in the range of an intron length) were fused to a common gene locus. A total of 34,438 gene loci were defined in this way (Table 1), half of which are represented by transcripts (17,158). The gene loci with corresponding transcripts have a mean length of 1951 bases, while those without transcript support have a mean length of only 579 bases. This suggests that the majority of unsupported gene loci may represent false positive predictions, which is fairly typical of highly fragmented genome assemblies.

There are scores of complex repeats in the genome, which contains a minimum of 484 integrase domains (PF00665) and 1014 reverse transcriptase domains (PF07727 and PF00078). This greatly exceeds the number present in other Amoebozoa species such as *A. castellanii* and *D. disoideum*, which contain only 38 and 118 reverse transcriptase domains, respectively. In total, *P. polycephalum* complex repetitive elements contribute approximately 3 Mb or 1.3 % to the size of the genome (Table S4). Most *P. polycephalum* genes contain multiple introns, many of which are composed of extended repeats flanked by splicing signals. The median size of the 160,988 introns confirmed by transcript data is 231 bases, and the total number of predicted introns is 676,387. Thus, the mean number of confirmed introns per gene is approximately 5. If only the genes with transcript data are counted, the number of introns per gene increases to well above 9. This intron frequency is higher than the estimated number of introns in the last common ancestor of eukaryotes (Zhou, et al. 2014). Thus, *P. polycephalum* has likely acquired introns during its species history.

Domain analysis

Using iprscan (Zdobnov and Apweiler 2001) we defined the domain content for each gene locus. A comparison with *A. castellanii* and *D. discoideum* showed that *P. polycephalum* has the highest number of domain hits of the three organisms, reflecting the larger genome and the higher gene count. In terms of distinct domains, *A. castellanii* and *P. polycephalum* are similar, whereas fewer types of domains have been identified in *D. discoideum,* suggesting domain loss in the social amoebae (Table 2). Further analysis of domains revealed that 304 domains were present at least once in the genomes of *P. polycephalum* and *A. castellanii*, but absent from *D. discoideum* (Table S2). Among these are a considerable number of domains associated with signaling functions. Domains enriched compared to those of *A. castellanii* and *D. discoideum* include transposon domains and, again, domains associated with signaling functions (Table S3).

Non-coding RNAs

Small non-coding RNAs perform important functions in all organisms, but can be difficult to annotate *de novo* due to their short length and their poorly-defined sequence patterns. However, data obtained in a previous mitochondrial RNA-Seq experiment allowed us to annotate a number of small non-coding *P. polycephalum* RNAs encoded in the nuclear genome. This resulted in the confirmation of many known small *P. polycephalum* RNAs (Supplemental Spreadsheet 1, 1st sheet). In addition, we found 24 novel transfer RNAs, the spliceosomal U2 and U6 RNAs, 30 box C/D snoRNAs, 9 novel lncRNAs, which clustered into three groups by sequence similarity, as well as 24 RNAs shorter than 200 nucleotides clustering into three groups and 28 singleton RNAs shorter than 200 nucleotides (Supplementary Spreadsheet 1, 2nd sheet). These latter RNAs and the box C/D snoRNAs have no detectable sequence similarity with any sequence in the non-redundant nucleotide database.

We ran tRNAscan-SE (PMID: 9023104) with default parameters on the genome, which resulted in 319 putative tRNA loci and 248 unique putative tRNA sequences (Supplementary Spreadsheet 1, 3rd sheet). These included all but one (tRNA-Tyr1 in Supplementary Spreadsheet 1, 2nd sheet) of the 56 new tRNAs identified from the sequencing experiment as well as all previously annotated *P. polycephalum* tRNAs with the exception of tRNA-Lys1, tRNA-Ser2, and tRNA-Asp1. A total of 201 putative tRNA loci with 156 unique putative tRNA sequences were novel. These included tRNAs for all 20 amino acids as well as 6 putative tRNA-SeC loci (each with a unique sequence) and four putative tRNA loci of undetermined type (each with a unique sequence).

PPR proteins

The *P. polycephalum* genome is particularly rich in genes encoding pentatricopeptide repeat (PPR) proteins. PPR proteins are sequence-specific RNA-binding proteins that act in diverse processes of RNA maturation, mainly in mitochondria and chloroplasts (Schmitz-Linneweber and Small 2008). The PPR motif is a loosely conserved 35 aa motif identifying individual ribonucleotides on a one-repeat-per-nucleotide basis (Barkan, et al. 2012). As in other taxa, the *P. polycephalum* proteins contain PPR arrays of variable sizes ranging from only a few up to more than 20 tandem PPRs in individual proteins. With some 100 PPR protein genes, *P. polycephalum* features much more than tenfold the number found in other slime mold genomes such as *Polysphondylium pallidum* or *D. discoideum*. In fact, the set of PPR proteins present in *P. polycephalum* is significantly larger than in other

Amoebozoa and, more broadly, within the Amorphea super-domain at large. Indeed, larger numbers of PPR proteins have hitherto only been observed in Dinoflagellates and in land plants, where large families of PPR proteins were initially discovered (Small and Peeters 2000). Two observations are particularly intriguing: In contrast to other Amorphea, *P. polycephalum* has a highly derived mitochondrial genome, the transcripts of which are affected by abundant and diverse types of RNA editing, including alterations of transcript sequences by nucleotide insertions and base conversions (Bundschuh, et al. 2011; Mahendran, et al. 1991; Takano, et al. 2001). We speculate that some of the PPR proteins in *P. polycephalum* take part in these processes. Most notably, 16 of the *P. polycephalum* PPR proteins are “plant-like” in carrying a carboxyterminal DYW domain with cytidine deaminase similarity (Fig. S21) (Schallenberg-Rüdinger, et al. 2013). Several such DYW-type PPR proteins have been characterized as site-specific C-to-U editing factors in plants. Intriguingly, the first discovery of PPR proteins with a DYW domain outside of land plants in the genome of *Naegleria gruberi* led to the subsequent discovery of C-to-U editing in the mitochondria of this heterolobosean protist (Knoop and Rüdinger 2010; Rüdinger, et al. 2011). As in plants and *Naegleria*, the residues likely coordinating a zinc ion for deaminase activity are highly conserved in the *P. polycephalum* DYW domains. Given that in its natural environment there is direct physical contact with the slime mold growing on decaying plant materials, horizontal gene transfer (HGT) from plants may be a possible source of the *P. polycephalum* PPR genes. However, no evidence for particularly high similarity of any *Physarum* PPR protein to a homologue in another taxon indicating a recent HGT has been found. The DYW-type PPR protein families of *Physarum polycephalum, Naegleria gruberi* and the moss *Physcomitrella patens,* for example, cluster taxon-wise without evidence for recent HGT (Fig. S21C).

Moreover, whereas plant DYW-type PPR proteins feature characteristic “PLS-type” PPR arrays with alternating long (L) and short (S) variants of the classic (P) PPRs, most *Physarum* DYW-type homologues display “LS”-type repeats (Fig. S21A).

Metabolism and the cytoskeleton

Pteridine metabolism

Pteridines comprise a group of molecules that contain pteridine (pyrimido [4,5-b] pyrazine), a bicyclic ring system, as a common structural element. Pteridines (e.g. folate, biopterin, riboflavin etc.) function as important cofactors (or their precursors) of various enzymatic reactions. Pterines are synthesized from the common precursor GTP (guanosine 5’ triphosphate).

Analysis of the *P. polycephalum* genome for pteridine metabolic enzymes revealed that it is unusual in encoding enzymes for all common pteridine biosynthetic pathways (for the biosynthesis of molybdopterin, tetrahydrofolate, tetrahydrobiopterin, 7-aminomethyl deazaguanine, and riboflavin) and, in addition, the enzymes alkylglycerol monooxygenase and nitric oxide synthases. Alkylglycerol monooxygenase and nitric oxide synthase, in its full length form with oxygenase and reductase domains, are regularly found only in animals. Nitric oxide synthases are required for sporulation in *P. polycephalum* by acting via cGMP signaling (Golderer, et al. 2001). This link between animal-type NO synthases and cyclic nucleotide signaling, which is widely used in *P. polycephalum* (see Section Signaling), is noteworthy and to date is unique in unicellular eukaryotes. This suggests an ancient origin of this typically animal-type signaling enzyme and its signal transduction mechanism via cGMP.

Another unique finding is that riboflavin biosynthetic enzymes are encoded as two tri-functional reading frames (Fig. S10) rather than by separate cistrons, a feature thus far not observed for sequences from any other species represented in Genbank.

P. polycephalum is much more diverse than *D. discoideum* in its pteridine metabolic equipment. It contains enzymes for the biosynthesis of 7-aminomethyl-7-deaza guanine (preQ1), a precursor for the special tRNA base queuosine (Q), as well as for the biosynthesis of riboflavin. The reading frames for nitric oxide synthases found in *P. polycephalum* do not occur in the *D. discoideum* genome. Furthermore, another surprising feature is the comparatively frequent presence of homologous enzymes that catalyze the same enzymatic reaction, whereas most species including *D. discoideum* use only one. These homologs share in the mean 72 % sequence similarity, have the motifs required for enzymatic activity conserved, are expressed and are therefore expected to be enzymatically active. This assumption has been experimentally confirmed for the two inducible nitric oxide synthase homologs NOS2a and NOS2b (Messner, et al. 2009). The reasons for or advantages of this apparent redundancy are currently unknown. On the other hand, *D. discoideum* produces an additional unique pteridine, dictyopterin, an isomer of biopterin (Klein, et al. 1990) that is missing in *P. polycephalum*.

For the 24 genes analyzed in Table S5 encoding for 26 enzymes, closest orthologs for twelve were found in *Amoebozoa* (eight of these in *Dictyosteliida*), six in bacteria, four in animals, three in *Capsaspora owczarzaki*, an *Opisthokonta* of uncertain placing, and one in red algae (Table S5).

Shikimate pathway

The *P. polycephalum* genome also contains genes encoding enzymes for complete shikimate and aromatic amino acid synthesis pathways, which are not present in either the social amoebae or Entamoebae (Richards, et al. 2006). In fungi, alveolates, and oomycetes a pentafunctional polypeptide is formed (AROM). This organization is also observed in *P. polycephalum* (gene locus 13138), strongly suggesting that this is the ancestral state of the shikimate pathway protein domain structure. In contrast, it was recently shown that *A. castellanii* has an unusual arrangement of such pathway genes (Henriquez, et al. 2015), which might represent a later rearrangement within the Amoebozoa clade. Like all other free living Amoebozoa so far, *P. polycephalum* contains a phosphoenolpyruvate carboxylase (gene_locus_38419), which is primarily found in bacteria and plants. The purpose of this enzyme could be to supply the TCA cycle with C4 bodies when they are required for amino acid biosynthesis.

Actin cytoskeleton and motor domains

The very intriguing oscillatory cytoplasmic streaming in *P. polycephalum* triggered extensive studies on the cytoskeleton. It was first reported by Noburo Kamiya (Kamiya 1940), *i.e.* long before actin and myosin had been identified as driving forces in muscle and non-muscle cells. At up to 1350 µm/sec, the rhythmic streaming in different developmental stages of *P. polycephalum* is among the fastest intracellular motilities known to date (Kamiya 1959; Wohlfarth-Bottermann 1979). For comparison: an average kinesin moves with a velocity of about 0.6 µm/sec along a microtubular track.

In the current study many major families of actin-binding proteins have been found (Supplemental Spreadsheet 2). The profilins as members of G-actin sequestering proteins, fragmin as representative of the Ca^{2+}-dependent F-actin severing proteins, the heterodimeric F-actin capping proteins (formerly *'Physarum* beta-actinin'), F-actin crosslinking proteins like alpha-actinin and filamin, myosins as major motor proteins, and formin-like proteins with FH1 and FH2 domains are present in comparable numbers as in *D. discoideum* and more complex organisms. Even peculiar actin isoforms like *D. discoideum* filactin, an actin with a large N-terminal extension (Joseph, et al. 2008) could be identified in the *P. polycephalum* genome.

A seemingly unique and exciting feature of the actin cytokskeleton in *P. polycephalum*, however, is the presence of the actin-fragmin kinase (Contig 47652). Fragmin, homologous to severin in *D. discoideum* and gelsolin in more complex organisms (Yin, et al. 1990), was first described by the Hatano group (Hasegawa, et al. 1980) and its 1:1 complex with actin was further investigated by Vandekerckhove and colleagues (Eichinger, et al. 1996; Gettemans, et al. 1992). The latter group isolated a kinase which phosphorylated actin, but only in the 1:1 actin-fragmin complex. Upon phosphorylation the activity of the actin-fragmin complex changed to an F-actin capping function and was, therefore, a putative regulator of fast changes in local viscoelasticities in the F-actin network, a prerequisite for cytoplasmic streaming. Surprisingly, the protein sequence of the purified protein did not show any of the well known protein kinase domains (Hanks and Hunter 1995) and it could not be excluded that a minor contamination was responsible for actin phosphorylation. Only the crystallization of the protein finally showed that it was in fact a kinase whose structure was in its catalytic domain nearly identical to protein kinase A (Steinbacher, et al. 1999). These data lead to major conclusions: The actin-fragmin kinase is clearly an example for convergent evolution. There is no common ancestor as in the case of divergent evolution of conventional protein kinases. Thus, the evolutionary pressure is on the structure of the molecule in determining its function. There is no need to keep a distinct amino acid sequence with a certain arrangement of protein domains, just structure/function relationships count. Consequently, this type of actin phosphorylation in *P. polycephalum* is possibly not unique as suggested above. One cannot exclude convergent evolution of kinases with similar substrate specificity also in other organisms as these molecules may have completely unrelated sequences and are not easily detectable by sequence similarities. Perhaps most importantly: this example demonstrates the limitations of any whole genome approach. In the end only biochemistry and molecular cell biology will cross the border from 'gene products of unknown function' to proteins with known activities.

P. polycephalum possesses a wealth of motor domains. We searched all defined gene loci for the presence of such domains. With this initial search we found 43 gene loci with a kinesin domain (PF00225), 23 gene loci with a dynein heavy chain domain (PF03028), and 53 gene loci with myosin motors (PF00063). Since the genome assembly is fragmented and thus gene loci likely do not always represent a whole gene, we further investigated these loci making use of the reference transcriptome library. This approach enabled us to define eight complete dyneins, one N-terminal half, and two N-terminal fragments. Ten further fragments of varying length show homology to the middle and C-terminal parts of dyneins. Thus, the *P. polycephalum* genome likely encodes at least eleven dynein proteins. This number is in the same range as in other species.

With the same approach we defined fifteen complete or nearly complete myosin proteins and sixteen fragments of varying lengths. Five of the full length genes and seven of the fragments have the highest similarity to class VII unconventional myosins indicating an amplification of this family in *P. polycephalum*.

Kinesins have a wider length range than the other motor proteins and the regions outside the motor domain are highly variable. This makes it impossible to distinguish between full length genes and fragments as with the other motor domain proteins. Matching possible kinesin gene loci with transcript data yielded 39 defined loci. Three of these are small without transcript data. The mean length of the others is well above 700 amino acids. Thus we conclude, that at least 36 kinesin genes reside in the *P. polycephalum* genome. Strikingly, most of the gene loci we defined here are represented in our reference transcriptome library indicating their expression.

Dynein domains and, much more pronounced, kinesin domains are overrepresented in *P. polycephalum* as compared to two other organisms within the Amoebozoa, *D. discoideum* and *A. castellanii* (Table S3). The enrichment of dynein domains is likely due to the fact that *P. polycephalum* possesses flagella. It was shown previously that a plasmodial species, *Reticulomyxa filosa*, has amplified its complement of kinesin domain bearing genes, which may be associated with the increased requirements of intracellular transport in a huge cell (Glöckner, et al. 2014). We see this gene family expansion also in *P. polycephalum,* hinting at a common prerequisite for acquiring a plasmodial life stage.

Cell cycle regulation, cellular signalling, and photoreceptors

The detection of external stimuli and the processing of these stimuli into an appropriate response is an essential property of all living organisms. For a free-living, motile, phagocytotic, unicellular organism like *P. polycephalum*, physiologically relevant stimuli may include environmental factors such as light, humidity, temperature, pH, osmolarity, mechanical stress, or chemical signals released by mates, prey, predators, competitors or symbionts*.* As a true free-living species *P. polycephalum* should have retained most, if not all, components of such signaling systems inherited from the last common ancestor (LCA) of Amoebozoa. Since Opisthokonta and Amoebozoa form together the Amorphea (Adl, et al. 2012), it might even be possible to trace inventions thought to have appeared only in Metazoa back to this LCA.

In eukaryotes including mammalian cells, much of the sensory input directly or indirectly feeds into the control of the mitotic or meiotic cell cycle as part of the sensory control of developmental decisions or programs. Therefore genes encoding potential cell cycle regulatory proteins were also characterized.

Key cell cycle regulators are conserved between Amoebozoa and Metazoa

Building on early, pioneering studies of eukaryotic cell cycle regulation in *P. polycephalum* (Rusch, et al. 1966; Sachsenmaier, et al. 1972; Solnica-Krezel, et al. 1991), the availability of its genome sequence now provides unique opportunities to study the functional dynamics of the regulation of mitosis, meiosis, and the coregulation of cell cycle and cell fate decision. By analysing macroscopic samples taken at regular time intervals from a single plasmodial cell at subsequent points of a time series one can exploit the remarkable natural mitotic synchrony of plasmodial nuclei (see Supplementary Results and Discussion) and resolve time-dependent regulatory events within individual cells at transcriptome and proteome-wide scales. In this context the current finding that the control system of the mitotic cycle of *P. polycephalum* is typical of most eukaryotes, including plants and animals while different to that of the yeasts (Fig. 2) makes an important point.

Much of our knowledge of the molecular mechanisms controlling progression through the eukaryotic cell cycle has come from studies of model organisms, notably Fungi (budding yeast and fission yeast) and Metazoa (frog and fruit fly embryos and mammalian cell lines). Although there are deep functional similarities across all these organisms, suggesting a “universal” control mechanism for eukaryotes (Nurse 1990; Nurse 2002), there are differences in the regulators that control the G1/S transition in yeast cells (“start”) and animal cells (“restriction point”). The key decision point for proliferation and development in animal cells occurs in G1 phase. Recent reviews have suggested that the metazoan regulatory proteins E2F, Rb and Cyclin D may constitute the primitive G1/S control system found in most eukaryotes (Cross, et al. 2011; Doonan and Kitsios 2009; Harashima, et al. 2013). In support of this hypothesis, key cell cycle regulators appear to be strikingly conserved across Amoebozoa including *P. polycephalum.* The ancestral G1/S regulators were seemingly displaced in a fungal ancestor of the yeasts by unrelated but functionally analogous proteins (SBF, Whi5 and Cln3) (Medina *et al.*, unpublished results) (Fig. 2). Thus, *P. polycephalum* offers an attractive model and alternative to yeasts for studying the functional dynamics of a mammalian cell type mitotic cycle.

G-protein coupled receptors and heterotrimeric G-proteins

The G-protein coupled receptors (GPCRs) are a large and diverse family of transmembrane proteins in metazoa and other eukaryotes that detect a broad range of physical and chemical stimuli. Based on sequence similarity GPCRs are subdivided into six families, where members of each family detect different sets of ligands. We found at least 146 GPCRs, with representatives from all families except family 4, the fungal pheromone receptors (Fig. S13). The largest number of GPCRs (42) are present in family 1, the rhodopsin-like receptors, which apart from light receptors also contains receptors for cytokines and neuropeptides. The number of GPCRs in *P. polycephalum* is considerably larger than that for other Amoebozoa (Table 3A), such as *D. discoideum* with 55 GPCRs (Heidel, et al. 2011) and *A. castellanii* with just 35 (Clarke, et al. 2013).

The number of heterotrimeric G-proteins, which are activated by GPCRs as the next step in the signal processing cascade, is also large with 26 G-alpha subunits (Fig. S14), as compared to 12 in *D. discoideum* and six in *A. castellanii.* Thus, these gene families have been considerably expanded, possibly to cope with a wealth of different environmental conditions.

Blue light photoreceptors and phytochromes

There are distinct classes of chromoproteins that act as photoreceptors in both prokaryotes and eukaryotes in sensing the visible part of the spectrum including near UV: cryptochromes, LOV-domain photoreceptors, rhodopsins, and phytochromes (Heintzen 2012). Photoreceptor domains that are light-sensitive through covalently or non-covalently bound chromophores are often attached to signaling domains that relay the photosensory output to the cellular signaling network. While photobiochemical and physiological functions of quite a number of photoreceptors are well-studied in bacteria, plants, fungi, and animals, this is not the case for the members of the amoebozoa clade.

At least two photoreceptors, a phytochrome-like and a blue light photoreceptor act synergistically in controlling sporulation in *P. polycephalum* in response to far-red and to blue light (Lamparter and Marwan 2001; Starostzik and Marwan 1995a; Starostzik and Marwan 1995b) (for details see Supplementary Methods, Results and Discussion), but these receptors had not been identified at the molecular level. Five members of three classes of putative photoreceptors are expressed by *P. polycephalum*: one cryptochrome and one photolyase, one LOV-domain photoreceptor, and two phytochromes (Fig. 3).

Cryptochrome and Photolyase. Phypoly-transcript_04617 (*cryA,* cryptochrome-like photoreceptor*; gene locus 1148*) encodes a protein with highest similarity to chicken cryptochrome-2 and other related animal cryptochromes. Downstream of the DNA-photolyase and FAD-binding domain of *cryA* is a short C-terminal part of approximately 100 amino acids encoded in the genomic sequence that is missing in Phypoly-transcript_04513 (gene locus 3553 and 3554) which shares greatest sequence similarity with DNA-photolyases and is designated as *plyA* (for photolyase A type photoreceptor) and is presumably a DNA-photolyase rather than a sensory blue light photoreceptor. The physiological role of both *plyA* and *cryA* in *P. polycephalum* however remains to be established.

LOV-domain photoreceptor. The Phypoly-transcript_01902 (*gene locus 13045*) encodes a protein with a domain architecture composed of RasGAP, PB1, PAS_9, SAM, and PB1 superfamily domains (Fig. 3). BLAST analysis revealed the highest similarity to the phototropin-2 photoreceptor from *Arabidopsis thaliana* (P93025) and to other LOV domain-carrying phototropins from plants and bacteria as well as to the fungal white collar 1 (WC1) blue light photoreceptor from *Neurospora*. The PAS_9 domain of the predicted *P. polycephalum* LovA protein contains the highly conserved NCRFLQ motif with the cysteine residue serving as a potential chromophore binding site embedded in additonal highly conserved residues (Fig. S18) that are involved in chromophore binding and crucial for the photochemical properties of phototropin-type photoreceptors (Heintzen 2012). The domain composition of the predicted protein and the highly conserved residues for chromophore binding in the PAS_9 domain suggests that LovA may act as an unconventional blue light photoreceptor that integrates, modulates, and/or relays multiple signals and might even bind DNA through its PAS_9 domain.

Phytochromes. Originally considered to be plant-specific photoreceptors, phytochromes regulate many aspects of metabolism, motility, gene regulation and development (Heintzen 2012). Even before prokaryotic genome projects uncovered the early evolutionary origin of phytochromes, phytochrome-like photoreceptors were discovered in *Aspergillus* (Mooney and Yager 1990) and *P. polycephalum* (Lamparter and Marwan 2001; Starostzik and Marwan 1995b). Two phytochrome genes, *phyA* and *phyB*, are expressed in *P. polycephalum*, partially encoded by transcripts 20261 (gene locus 28349) and 03416 (gene locus 5996), respectively. Although the sequence identity between PhyA and PhyB is only 33.8 %, the two proteins share a similar domain architecture with the phytochrome-type PAS_2, GAF, PHY domain arrangement at the N-terminus and a C-terminal hybrid kinase-like part, composed of HisK, HATPase, and REC domains (Fig. 3). A BLAST search of the PAS_2, GAF, PHY domain portion of the proteins against the UniProt database revealed closest similarity to *Nostoc* PHYA (Q9LCC2) and four other bacterial phytochromes followed by plant phytochromes. Sequence alignment with the most similar phytochromes suggests that the chromophore binding site may be a cysteine close to the N-terminus that is conserved between *P. polycephalum* PhyA and PhyB and various bacterial phytochromes, while the cysteine of the CHxxYxxNMG motif that serves as a chromophore binding site in plant phytochromes is replaced by valine in the two *P. polycephalum* proteins (Fig. S20). Whether the photochemistry of the two phytochromes is identical or one of them may be specifically synthesized in the dark in its far-red light absorbing P_{fr} form (Lamparter and Marwan 2001) to trigger sporulation (see Supplementary Information) remains to be established. The deletion upstream of the conserved PASDIPPQARRL motif in PhyA as compared to PhyB could conceivably cause an according functional difference.

Sensor histidine kinases/phosphatases

Other important receptors for external stimuli are the sensor histidine kinases/phosphatases (SHKPs), which are very abundant in prokaryotes and also present in significant numbers in plants, fungi and Amoebozoa, but not in metazoa (Wolanin, et al. 2002). Sensor histidine kinases/phosphatases initiate forward or reverse relay, respectively, of a phosphoryl group to/from a conserved aspartate in a response regulator. The response regulator subsequently regulates the activity of an effector, such as an enzyme or transcription factor. The *P. polycephalum* genome contains ~51 SHKPs with a large variety of different functional domain architectures (Fig. 4A). Many of these SHKPs contain small molecule or light sensing domains, such as the GAF, PAS/PAC, or phytochrome domains, which are likely to detect stimuli and activate phosphotransfer, while others contain protein kinase or AAA-ATPase domains. These domains could be downstream effectors, which are regulated by phosphorelay. The number of SHKPs in *P. polycephalum* is about equal to that in *A. castellanii* and three times larger than that in *D. discoideum* (Table 3B) but there are up to five times more response regulators in *P. polycephalum* than in either *D. discoideum* or *A. castellanii*.

While most *P. polycephalum* response regulators have no additional features, two have leucine-rich repeats and two have a phosphodiesterase domain and are closely related to the *D. discoideum* cAMP phosphodiesterase RegA (Fig. 4B). In *Dictyostelids* RegA is a negative regulator of encystation and sporulation, which both require activation of PKA by cAMP (Du, et al. 2014; Shaulsky, et al. 1998). In *P. polycephalum,* sporulation is induced by light and this induction was shown to be mediated by a phytochrome-type receptor (Starostzik and Marwan 1995b). The light-sensing part of this photoreceptor is likely to be the phytochrome domain that is present in two *P. polycephalum* SHKPs (see below; Fig. 4A), which suggests that the stimuli that control sporulation in *P. polycephalum* also act on intracellular cAMP levels by regulating RegA activity.

Cyclic nucleotide signalling

The cyclic nucleotides cAMP, and to a lesser extent cGMP, are widely used as intracellular second messengers for a broad range of external stimuli in all domains of life. cAMP is particularly important for Dictyostelid development, where it regulates encystation, progression through multicellular development, maturation of spore and stalk cells, and maintenance of spore and cyst dormancy by acting on PKA (Schaap 2011). In addition, Dictyostelids also use cAMP extracellularly as a chemoattractant to organize fruiting body morphogenesis in all species, and aggregation in a subset that contains the model organism *D. discoideum* (Alvarez-Curto, et al. 2005). In contrast, there are a few documented roles for cAMP or cGMP in *P. polycephalum*. cGMP plays a role in the induction of sporulation in *P. polycephalum* (Golderer et al., 2001) and there is sporadic evidence of roles for cAMP in cell division, motility and gravity sensing (Block, et al. 1998; Kuehn 1972; Matveeva, et al. 2012).

To evaluate the prevalence of cAMP and cGMP signaling in *P. polycephalum*, we investigated the presence of the cyclases, binding proteins, and phosphodiesterases that respectively synthesize, detect, or degrade cyclic nucleotides.

To our surprise the number of cyclase and cyclic nucleotide binding proteins in *P. polycephalum* by far surpasses that in *D. discoideum.* Firstly, *P. polycephalum* has 64 nucleotidyl cyclases against 5 in *D. discoideum*. *A. castellanii* has even more (67), but 66 of these cyclases belong to a massively amplified family of transmembrane proteins that contain a cyclase domain, flanked by two protein kinase domains. *P. polycephalum* has a family of 43 of those proteins, suggesting a common requirement for amplification of these cyclases in some Amoebozoa. *P. polycephalum* has in addition 21 cyclases with a highly variable functional domain architecture (Fig. S16). This set is dominated by proteins with 2 sets of 6 transmembrane domains that flank 2 cyclase domains. This is the configuration of most mammalian adenylate cyclases and of *D. discoideum* ACA and GCA. In addition there are cyclases with GAF and PAS/PAC domains and with domains involved in either actin remodeling, ubiquitination or calcium export. *P. polycephalum* also has a close homolog of *D. discoideum* SGC, a guanylate cyclase that is involved in chemotaxis, and it has a close homolog of *D. discoideum* and *A. castellanii* AcrA, an adenylate cyclase that is essential for spore maturation in *D. discoideum* (Loomis 2014).

The number of cyclic nucleotide phosphodiesterases in *P. polycephalum* is about equal to that in *D. discoideum* and *A. castellanii* (Fig. S15). However, the number of proteins predicted to be capable of binding cAMP or cGMP is truly astonishing and exceeds the numbers in *D. discoideum* and *A. castellanii* by five and four fold, respectively (Table 3). While the majority of these proteins only have cNMP binding domain(s), there are also homologs of the *D. discoideum* cGMP binding proteins GbpC and GbpD, which additionally have protein kinase, RasGEF and DEP domains (Fig. S17). There is also a homolog of *D. discoideum* PdeE, in which an intrinsic Lactamase_B domain acts as a cAMP phosphodiesterase. In addition, there are putative cNMP binding proteins with additional RhoGEF, VWA and P2X receptor domains that are not present in other organisms. The multitude and variety of cNMP binding domains and nucleotidyl cyclases in *P. polycephalum* indicates that cyclic nucleotide signalling is likely to play a very dominant role in its physiology and development.

Protein kinases

The protein kinases that modify the function of downstream effector proteins by phosphorylation of serine/threonine (S/T) or tyrosine (Y) residues are the most common intermediates of signal processing cascades. Humans have 518 protein kinases and even yeast has over a hundred (Hanks 2003). Inspection of the *P. polycephalum* transcriptome revealed the presence of a total of 447 proteins with S/T, Y or S/T/Y (dual specificity) kinase domains (see Supplemental Methods, Results, and Discussion and Table 3 for further details). This is 1.5 and 1.2 fold more than in *D. discoideum* and *A. castellanii*, respectively, and therefore a more modest increase than observed for the other signal transduction proteins described above.

The initial analysis based on Interpro identifiers for substrate specificity revealed the presence of 29 tyrosine kinases, but a more in depth comparison with sequence signatures of validated tyrosine kinases revealed that only four of those contain all essential residues for tyrosine substrate recognition (Fig. 5). Three of these proteins contain a single transmembrane domain that provides them with structural similarity to metazoan receptor tyrosine kinases (RTKs), which are the targets for a broad range of peptide growth factors, hormones and cytokines. In metazoa, tyrosine phosporylated substrates become binding sites for SH2 domain proteins, which further transduce the response (Liu and Nash 2012). *P. polycephalum* has 18 SH2 domain proteins (Fig. 6), of which three are similar to *D. discoideum* and metazoan STAT transcription factors, while *D. discoideum* and *A. castellanii* have 15 and 48 SH2 domain proteins, respectively.

Tyrosine kinases play very dominant roles in metazoan embryogenesis and adult physiology, and were until recently considered to be a hallmark of metazoan evolution because they were present only in metazoa and one of its unicellular allies, a choanoflagellate (King and Carroll 2001) and absent from yeast, plants and *D. discoideum* (Lim and Pawson 2010). This notion was overturned by the identification of some tyrosine kinases outside of Opisthokonta (Suga, et al. 2012) and the presence of 21 tyrosine kinases in *A. castellanii* (Clarke, et al. 2013), a member of Lobosa, one of the two major subdivisions of Amoebozoa. *P. polycephalum* and *D. discoideum* reside in the other subdivision Conosa (Schilde and Schaap 2013). The presence of tyrosine kinases in both subdivisions of Amoebozoa, in their sister group Opisthokonta, and in other eukaryote divisions indicates that tyrosine kinases were likely present in the last common ancestor to all eukaryotes and were selectively lost from many phyla. The more widely distributed tyrosine kinase like enzymes, e.g. *D.*

discoideum Pyk2, which can also phosphorylate SH2 domain tyrosines (Araki, et al. 2014), has probably taken over their role in these phyla.

Conclusions

P. polycephalum has for long been a classic model organism in cell biology. The accessibility of both genome sequence and extensive transcriptome data sets strongly enhances its usefulness as a model system. The availability of the transcriptome of amoebae and plasmodia at different stages of development now permits reverse genetic approaches through identifying mutants in genes of interest in mutant libraries obtained by chemical mutagenesis of amoebal cells. In addition, the transcriptome sequences provide the basis for quantitative proteomic approaches by mass spectrometry. Our analysis highlights a wealth of interesting molecular features and will help in enabling work at the molecular level in combining the well-established classic genetics with second generation sequencing approaches.

The results of the *P. polycephalum* sequencing project described here provide interesting information with respect to the evolution of signaling systems in the Amorphea branch of the eukaryotes. Virtually all aspects discussed here characterize *P. polycephalum* as an organism with higher molecular complexity than other sequenced Amoebozoa. This becomes most obvious for cellular signalling through the extensive use of two component signaling systems in parallel with receptor tyrosine kinase signal transduction. Other important aspects such as cell cycle regulation, cytoskeletal motor proteins, and the enzymes of the pteridine metabolism point in the same direction. Together with recent findings on *A. castellanii* (Clarke, et al. 2013), our results indicate that the molecular evolution of these features, notably signaling through receptor tyrosine kinases which has previously been considered a hallmark of multicellularity of animals, is deeply rooted in the Amorphea and have been secondarily lost in other amoebozoa like *D. discoideum* and in the fungi. Another interesting feature of *P. polycephalum* is the occurrence of photoreceptors: two bacterial-type phytochromes, not found in other Amoebozoa, a phototropin-like LOV domain blue light photoreceptor, and a cryptochrome. In conclusion, among the unicellular genetic model organisms *P. polycephalum* displays many features of animal cells and, with respect to its molecular complexity, the cross-talk of signaling molecules, and the resulting dynamic behavior. As such, it is anticipated that it will serve as an interesting model system contributing complementary information for the study of mammalian and other animal cells. In combination with its ability to form multinucleate giant plasmodia, this provides unique options for investigating fundamental biological processes in individual cells.

Supplementary Material

Supplementary File 1: Supplementary Methods, Results and Discussion

Supplementary File 2: Supplementary Spreadsheets

Acknowledgments

Funding for the sequence characterization of the *P. polycephalum* genome was provided by National Institutes of Health (NIH) grant HG003079 to RKW, The Genome Institute at Washington University School of Medicine. PS received funding from the Wellcome Trust and the BBSRC. IB was supported by the Excellence Initiative of Saxony-Anhalt through the International Max Planck Research School, Magdeburg. WM received funding from the German Federal Ministry of Education and Research (BMBF) via the FORSYS program. Pteridine work was funded by the Austrian Science Fund (FWF), P22406 to ERW. JMG was funded through (NIH GM54663, NSF MCB-1243687).

Literature Cited

Adl SM, et al. 2012. The revised classification of eukaryotes. Journal of Eukaryotic Microbiology 59: 429-493.

Aldrich HC, Daniel JW. 1982. Cell Biology of *Physarum* and *Didymium*. In. New York: Adademic Press.

Alvarez-Curto E, et al. 2005. Evolutionary origin of cAMP-based chemoattraction in the social amoebae. Proc. Natl. Acad. Sci. USA 102: 6385-6390.

Araki T, et al. 2014. Two *Dictyostelium* tyrosine kinase-like kinases function in parallel, stress-induced STAT activation pathways. Mol Biol Cell 25: 3222-3233.

Barkan A, et al. 2012. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS genetics 8: e1002910.

Block I, Rabien H, Ivanova K 1998. Involvement of the second messenger cAMP in gravity-signal transduction in *Physarum*. Adv Space Res 21: 1311-1314.

Braund E, Miranda ER editors. 9th Conference on Interdisciplinary Musicology – CIM14. papers2://publication/uuid/AC77D621-D506-4D95-A026-AC65533A364F; 2014 Mar 12: Berlin.

Bundschuh R, Altmüller J, Becker C, Nürnberg P, Gott JM 2011. Complete characterization of the edited transcriptome of the mitochondrion of *Physarum polycephalum* using deep sequencing of RNA. Nucleic Acids Research 39: 6044-6055.

Burland TG, Solnica-Krezel L, Bailey J, Cunningham DB, Dove WF 1993. Patterns of inheritance, development and the mitotic cycle in the protist *Physarum polycephalum*. Adv. Microb. Physiol. 35: 1-69.

Cavalier-Smith T 2003. Protist phylogeny and the high-level classification of Protozoa. European Journal of Protistology 39: 338-348.

Clarke M, et al. 2013. Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. Genome Biol 14: R11.

Cross FR, Buchler NE, Skotheim JM 2011. Evolution of networks and sequences in eukaryotic cell cycle control. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 366: 3532-3544.

Dee J 1987. Genes and development in *Physarum*. Trends Genet. 3: 208-213.

Dee J. 1982. Genetics of *Physarum polycephalum*. In: Aldrich HC, Daniel JW, editors. Cell Biology of *Physarum* and *Didymium*: Academic Press. p. 211-251.

Doonan JH, Kitsios G 2009. Functional evolution of cyclin-dependent kinases. Molecular Biotechnology 42: 14-29.

Dove WF, Dee J, Hatano S, Haugli FB, Wohlfarth-Bottermann K-E. 1986. The Molecular Biology of Physarum polycephalum. In. New York: Plenum Press. p. 253-269.

Du Q, et al. 2014. The cyclic AMP phosphodiesterase RegA critically regulates encystation in social and pathogenic amoebas. Cellular Signalling 26: 453-459.

Edgar RC 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics (Oxford, England) 26: 2460-2461.

Eichinger L, Bomblies L, Vandekerckhove J, Schleicher M, Gettemans J 1996. A novel type of protein kinase phosphorylates actin in the actin–fragmin complex. EMBO J. 15: 5547–5556.

Eichinger L, et al. 2005. The genome of the social amoeba *Dictyostelium discoideum*. Nature 435: 43-57.

Gettemans J, De Ville Y, Vandekerckhove J, Waelkens E 1992. *Physarum* actin is phosphorylated as the actin–fragmin complex at residues Thr203 and Thr202 by a specific 80 kDa kinase. EMBO J. 11: 3185–3191.

Glöckner G, Golderer G, Werner-Felmayer G, Meyer S, Marwan W 2008. A first glimpse at the transcriptome of *Physarum polycephalum*. BMC Genomics 9: 6.

Glöckner G, et al. 2014. The genome of the foraminiferan *Reticulomyxa filosa*. Curr Biol 24: 11 - 18.

Golderer G, Werner ER, Leitner S, Gröbner P, Werner-Felmayer G 2001. Nitric oxide synthase is induced in sporulation of *Physarum polycephalum*. Genes Devel. 15: 1299-1309.

Haindl M, Holler E 2005. Use of the giant multinucleate plasmodium of *Physarum polycephalum* to study RNA interference in the myxomycete. Analytical biochemistry 342: 194-199.

Hanks SK 2003. Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. Genome Biology 4: 111.

Hanks SK, Hunter T 1995. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J. 9: 576–596.

Harashima H, Dissmeyer N, Schnittger A 2013. Cell cycle control across the eukaryotic kingdom. Trends in Cell Biology 23: 345-356.

Hasegawa T, Takahashi S, Hayashi H, Hatano S 1980. Fragmin: a calcium ion sensitive regulatory factor on the formation of actin filaments. Biochemistry 19: 2677-2683.

Heidel A, et al. 2011. Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. Genome Res.: 1882-1891.

Heintzen C 2012. Plant and fungal photopigments. Wiley Interdisciplinary Reviews: Membrane Transport and Signaling 1: 411-432.

Henriquez FL, et al. 2015. The *Acanthamoeba* Shikimate Pathway has a Unique Molecular Arrangement and is Essential for Aromatic Amino Acid Biosynthesis. Protist 166: 93-105.

Itoh K, et al. 2011. DNA packaging proteins Glom and Glom2 coordinately organize the mitochondrial nucleoid of *Physarum polycephalum*. Mitochondrion 11: 575-586.

Joseph J, et al. 2008. The actinome of *Dictyostelium discoideum* in comparison to actins and actin-related proteins from other organisms. PLoS ONE 3: e2654.

Kamiya N 1940. The control of protoplasmic streaming. Science (New York, NY) 92: 462-463.

Kamiya N. 1959. Protoplasmic streaming. Wien: Springer-Verlag.

Kent JW 2002. BLAT – The BLAST like alignment tool. Genome Research 12: 656-664.

King N, Carroll S 2001. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. Proc Natl Acad Sci USA 98: 15032-15037.

Klein R, Thiery R, Tatischeff I 1990. Dictyopterin, 6-(D-threo-1,2-dihydroxypropyl)-pterin, a new natural isomer of L-biopterin. Isolation from vegetative cells of *Dictyostelium discoideum* and identification. Eur J Biochem 187: 665-669.

Knoop V, Rüdinger M 2010. DYW-type PPR proteins in a heterolobosean protist: plant RNA editing factors involved in an ancient horizontal gene transfer? FEBS letters 584: 4287-4291.

Kuehn GD 1972. Cell cycle variation in cyclic adenosine 3',5'-monophosphate-dependent inhibition of protein kinase from *Physarum polycephalum*. Biochem Biophys Res Commun 49: 414-419.

Lamparter T, Marwan W 2001. Spectroscopic detection of a phytochrome-like photoreceptor in the Myxomycete *Physarum polycephalum* and the kinetic mechanism for the photocontrol of sporulation by P_{fr}. Photochem. Photobiol. 73: 697-702.

Lim WA, Pawson T 2010. Phosphotyrosine signaling: evolving a new cellular communication system. Cell 142: 661-667.

Liu BA, Nash PD 2012. Evolution of SH2 domains and phosphotyrosine signalling networks. Philosophical Transactions of the Royal Society B: Biological Sciences 367: 2556-2573.

Loomis WF 2014. Cell signaling during development of *Dictyostelium*. Developmental Biology 391: 1-16.

Mahendran R, Spottswood MR, Miller DL 1991. RNA editing by cytidine insertion in mitochondria of *Physarum polycephalum*. Nature 349: 434-438.

Marchler-Bauer A, et al. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Research 43.

Matveeva NB, Teplov VA, Nezvetskii AR, Orlova TG, Beilina SI 2012. Involvement of cyclic adenosine monophosphate in the control of motile behavior of *Physarum polycephalum* plasmodium. Biofizika 57: 832-839.

Messner S, et al. 2009. *Physarum* nitric oxide synthases: Genomic structures and enzymology of recombinant proteins. Biochem. J. 418: 691-700.

Mooney JL, Yager LN 1990. Light is required for conidiation in *Aspergillus nidulans*. Genes & Development 4: 1473-1482.

Nurse P 1990. Universal control mechanism regulating onset of M-phase. Nature 344: 503-508.

Nurse PM 2002. Nobel Lecture: Cyclin dependent kinases and cell cycle control. Bioscience Reports 22: 487-499.

Richards TA, et al. 2006. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. Eukaryot Cell 5: 1517 - 1531.

Rüdinger M, Fritz-Laylin L, Polsakiewicz M, Knoop V 2011. Plant-type mitochondrial RNA editing in the protist *Naegleria gruberi*. RNA (New York, N.Y.) 17: 2058-2062.

Rusch HP, Sachsenmaier W, Behrens K, Gruter V 1966. Synchronization of Mitosis by the Fusion of the Plasmodia of *Physarum polycephalum*. The Journal of cell biology 31: 204-209.

Sachsenmaier W, Remy U, Plattner-Schobel R 1972. Initiation of synchronous mitosis in *Physarum polycephalum*. Exp. Cell Res. 73: 41-48.

Sauer HW. 1982. Developmental biology of *Physarum*. In. Cambridge: Cambridge University Press

Schaap P 2011. Evolutionary crossroads in developmental biology: *Dictyostelium discoideum*. Development 138: 387-396.

Schallenberg-Rüdinger M, Lenz H, Polsakiewicz M, Gott JM, Knoop V 2013. A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. RNA biology 10: 1549-1556.

Schedl T, Dove WF 1982. Mendelian analysis of the organization of actin sequences in *Physarum polycephalum*. Journal of Molecular Biology 160: 41-57.

Schilde C, Schaap P 2013. The Amoebozoa. Methods Mol Biol 983: 1-15.

Schmitz-Linneweber C, Small I 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends in Plant Science 13: 663-670.

Shaulsky G, Fuller D, Loomis WF 1998. A cAMP-phosphodiesterase controls PKA-dependent differentiation. Development 125: 691-699.

Small ID, Peeters N 2000. The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends in Biochemical Sciences 25: 46-47.

Solnica-Krezel L, Burland TG, Dove WF 1991. Variable pathways for developmental changes of mitosis and cytokinesis in *Physarum polycephalum*. The Journal of cell biology 113: 591-604.

Stanke M, Steinkamp R, Waack S, Morgenstern B 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res 32: W309-312.

Starostzik C, Marwan W 1995a. Functional mapping of the branched signal transduction pathway that controls sporulation in *Physarum polycephalum*. Photochem. Photobiol. 62: 930-933.

Starostzik C, Marwan W 1995b. A photoreceptor with characteristics of phytochrome triggers sporulation in the true slime mould *Physarum polycephalum*. FEBS Lett 370: 146-148.

Steinbacher S, et al. 1999. The crystal structure of the *Physarum polycephalum* actin-fragmin kinase: an atypical protein kinase with a specialized substrate-binding domain. EMBO J 18: 2923-2929.

Suga H, et al. 2012. Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases. Sci Signal 5: ra35.

Takano H, et al. 2001. The complete DNA sequence of the mitochondrial genome of *Physarum polycephalum*. Mol. Gen. Genet. 264: 539-545.

Tero A, et al. 2010. Rules for biologically inspired adaptive network design. Science (New York, NY) 327: 439-442.

Tsuda S, Zauner K-P, Gunji Y-P 2007. Robot control with biological cells. Bio Systems 87: 215-223.

Wohlfarth-Bottermann KE 1979. Oscillatory contraction activity in *Physarum*. J. Exp. Biol. 81: 15-32.

Wolanin P, Thomason P, Stock J 2002. Histidine protein kinases: key signal transducers outside the animal kingdom. Genome Biology 3: reviews3013.3011 - reviews3013.3018.

Yin H, Janmey P, Schleicher M 1990. Severin is a gelsolin prototype. FEBS Lett. 264: 78-80.

Zdobnov EM, Apweiler R 2001. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847-848.

Zhou K, Kuo A, Grigoriev IV 2014. Reverse transcriptase and intron number evolution. Stem Cell Investigation 1.

Tables

Table 1. Assembly, transcriptome, and gene prediction

Scaffold Assembly size (Mb)	188.75
Number of scaffolds	55119
Mean scaffold length	3424
N50 scaffold length	65980
Contig size (Mb)	173.6
Number of contigs	138064
Mean contig length	1256
N50 contig length	2197
Transcriptome contigs	733443
Reference transcriptome after clustering	31770
predicted gene models	47779
predicted genes without transcript support longer than 40 aa (all)	17199 (17280)
all gene loci (overlapping reference transcripts defined as one gene locus)	34438

Table 2. Domain distribution in three Amoebozoa

	P. polycephalum	*D. discoideum*	*A. castellanii*
All domains	19611	11307	17560
Different domains	3504	3192	3542

Table 3. Signalling proteins in amoebozoa.

Class of signalling proteins	*Pp*	*Dd*	*Ac*
A) G-protein coupled receptors			
GPCR family			
1 Rhodopsin-like	42	0	9
2 Secretin-like	23	1	5
3 metabotropic glutamate	38	17	0
4 fungal pheromone	0	0	0
5 frizzled-like	28	25	16
6 cAR-like	14	13	5
All	146	55	35
G-protein subunits			
alpha	26	12	6
beta	1	1	n.d.
gamma	1	1	n.d.
B) Sensor histidine kinases/phosphatases and response regulators			
SHKPs	51	16	48
Response regulators	27*)	5	5
C) Cyclic nucleotide signalling			
cNMP synthesis	64	5	67
cyclase kinases	43	0	66
cyclases (other)	21	5	1
cNMP detection	28	5	7
cNMP hydrolysis	11	8	10
D) Protein kinases			
Total	447	295	377
Tyrosine kinases	4	0	21
SH2 domain proteins	18	15	48

A) G-protein coupled receptors. Number of members in each of the six families of G-protein coupled receptors (GPCRs) and of subunits of heterotrimeric G-proteins in *Physarum polycephalum* (*Pp*), *Dictyostelium discoideum* (*Dd*) and *Acanthamoeba castellani* (*Ac*). See Fig. S13 and Fig. S14 for phylogenetic relationships between the *Physarum* proteins.

B) Sensor histidine kinases/phosphatases and response regulators. *) Note that the number of response regulators could be somewhat inflated by inclusion of sequence fragments that contain response regulators from incompletely assembled SHKPs. See Fig. 4 for phylogenies and functional domain architectures of the proteins.

C) Cyclic nucleotide signalling. Abundance of nucleotidyl cyclases, cAMP or GMP binding proteins and cyclic nucleotide phosphodiesterases. See Fig. S15 and Fig. S16 for phylogenies and functional domain architectures of cyclases and phosphodiesterases.

D) Protein kinases. Proteins with S/T, S/T/Y or Y protein kinase domains were retrieved from an Interproscan of all transcribed coding regions by the presence of the Interpro IPR002290, IPR008271, IPR001245, IPR020635 and/or IPR008266 domains. A total of 29 proteins contained the IPR020635, IPR008266 identifiers for Y protein kinases, but only 4 of those showed full consensus with validated tyrosine kinase specific sequences (see Fig. 5).

Figure Legends

Figure 1. *P. polycephalum*, life cycle and relationship to other Amoebozoa. A) Schematic representation of the different stages that form during the heterothallic or the apogamic life cycle as represented by the outer or the inner circle, respectively. One haploid (n), mononucleate amoeba hatches from each germinating haploid, mononucleate spore. Amoebae have four developmental options. They can propagate by cell division through an open mitosis or differentiate into a flagellate, a cyst, or a multinucleate plasmodium. **Heterothallic cycle:** Two amoebae of different mating type mate with each other to form a mononucleate, diploid (2n) zygote by karyogamy. The zygote develops into a diploid, multinucleate plasmodium. In the plasmodium, nuclei divide synchronously while the nuclear envelopes remain intact (closed mitosis). The plasmodium keeps growing as long as environmental conditions are favourable. A mature, multinucleate plasmodium can develop into a sclerotium to survive drought. Sporulation of a starving plasmodium is induced by visible light or heat shock. During sporulation, the protoplasmic mass develops into multiple fruiting bodies. Each sporangium contains hundereds of haploid (n), mononucleate spores that have been formed through meiosis. **Apogamic cycle:** A haploid, mononucleate amoeba that carries a *gadAh* mutant allele may develop into a multinucleate, haploid plasmodium without mating. Upon sporulation the low number of diploid nuclei that have been formed within the plasmodium give rise to viable spores. Apogamic development can be suppressed experimentally by elevated temperature which allows propagation of amoebal clones or the formation of a diploid plasmodium by mating of two amoebae of different mating type. As mating of

amoebal cells and plasmodium formation are controlled by different mating type genes, sophisticated approaches based on Mendelian genetics are possible (Dee 1987). B) Phylogenetic subtree of some Amoebozoa species with completed genomes. The Tree is based on 30 highly conserved genes and rooted with other eukaryotes. The complete tree including representatives from (plants, animals, and fungi) is shown in Fig. S22. The maximum likelihood method with the JTT matrix was used. *Entamoeba histolytica* genes were omitted for calculating the tree because endoparasites are known to evolve at different rates as compared to free-living organisms. For details see Supplemental Methods, Results, and Discussion.

Figure 2. Analysis of key cell cycle regulators across diverse eukaryotes. We searched each genome for cell division kinases (Cdk), cyclins (Cyc), G2/M regulators (Cdc25, Wee1), APC regulators (Cdc20, Fzr1), G1/S transcription factors (E2F/DP), and G1/S inhibitors (Rb); see Methods for details. Each entry is a conservative estimate of the number of family members in each genome. Top and bottom grey rows are common cell cycle gene names from *Homo sapiens* and *Arabidopsis thaliana*, respectively. Sub-families or alternative names of cell cycle genes are listed in parentheses. Metazoa (e.g. animals) and Fungi (e.g. yeasts) are members of the Opisthokonta (red), whereas Viridiplantae (e.g. land plants) are Archaeaplastida (green). *P. polycephalum* is a member of the Amoebozoa (grey), which are a sister group to the Opisthokonta. The last common ancestor of Amoebozoa and Opisthokonta is known as Amorphea (Adl, et al. 2012).

Figure 3. Domain architecture of the predicted photoreceptors encoded in the *P. polycephalum* transcriptome as determined by searching the NCBI Conserved Domain Database (Marchler-Bauer, et al. 2015). A description of each domain can be retrived from the database by searching for the name as it is displayed in this figure. See Supplementary Methods, Results and Discussion for predicted chromophore binding sites.

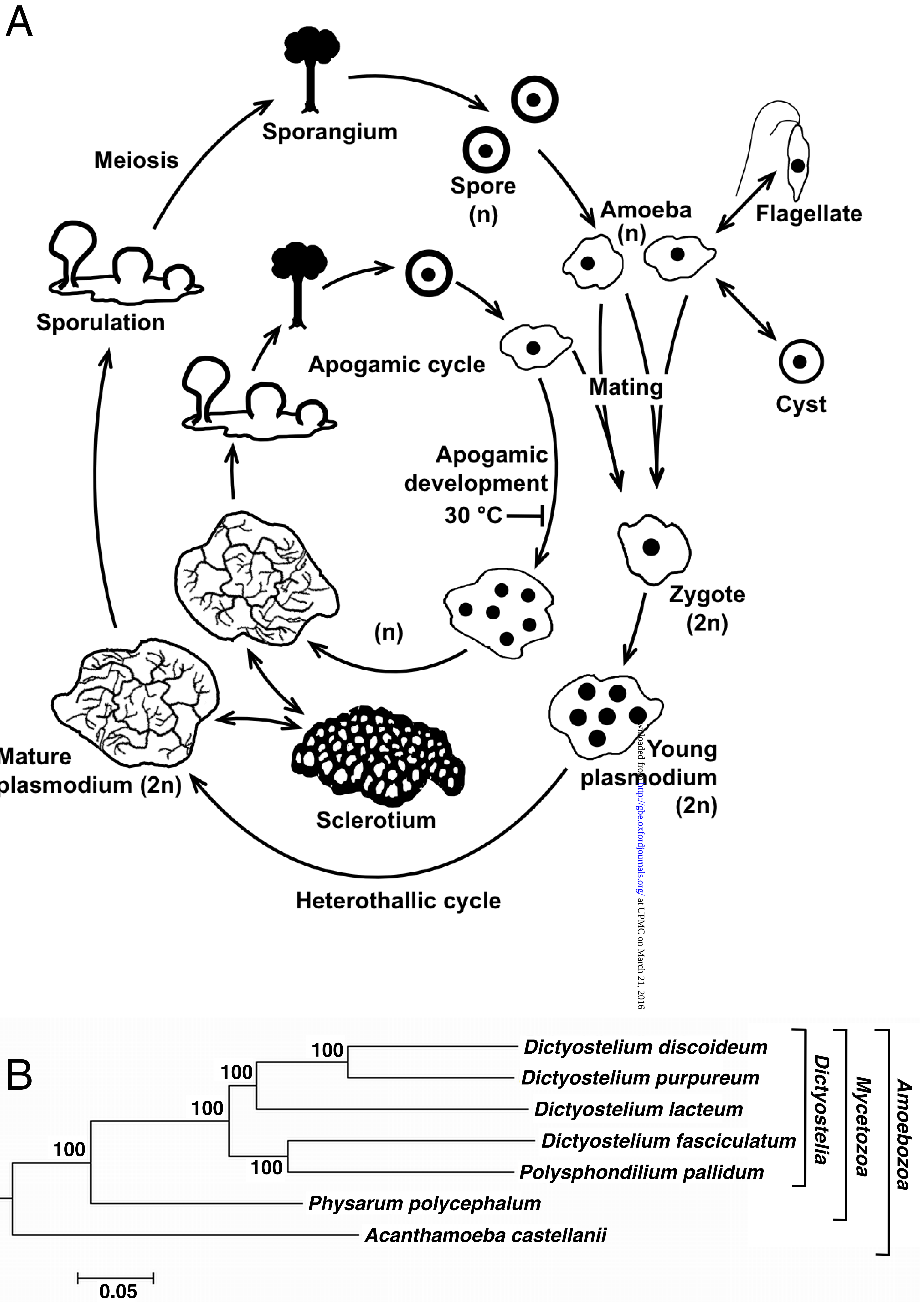
Figure 4. Sensor histidine kinases/phosphatases and phosphorelay receivers. *A. Sensor histidine kinases/phosphatase.* The *P. polycephalum* genome was queried by BLAST with all 15 Dictyostelid SHKPs, plant and fungal SHKPs and with the most divergent prokaryote SHKPs, while transcribed proteins were queried for the presence of Interpro domains IPR003661 and IPR003594 for the HisKA and HATPase-c moieties of SHKPs, respectively. The retrieved sets largely overlapped, but the transcriptome (Phypo identifiers in the tree) contained both the largest number and more complete SHKPs. Sequences were first aligned with all their closest relatives in other organisms, as identified by BlastP of the Genbank non-redundant sequences. However, in a phylogeny constructed from this alignment, most of the outgroup sequences clustered together. Their contribution therefore was reduced to the ones with greatest identity to individual clusters of *P. polycephalum* SHKPs. A second phylogeny was constructed from these outgroup and *P. polycephalum* sequences, using only segments encompassing the HisKA, HATPase-c and receiver domains, which are common to all or most retrieved SHKPs. The tree was annotated with the functional domain architecture of the proteins. The methods used for sequence alignment, alignment editing and Bayesian phylogenetic inference are the same as described in the legend to Fig. S13. Protein tags are colour-coded to

reflect the species of origin as shown, and Bayesian posterior probabilities of tree nodes are represented by coloured dots.

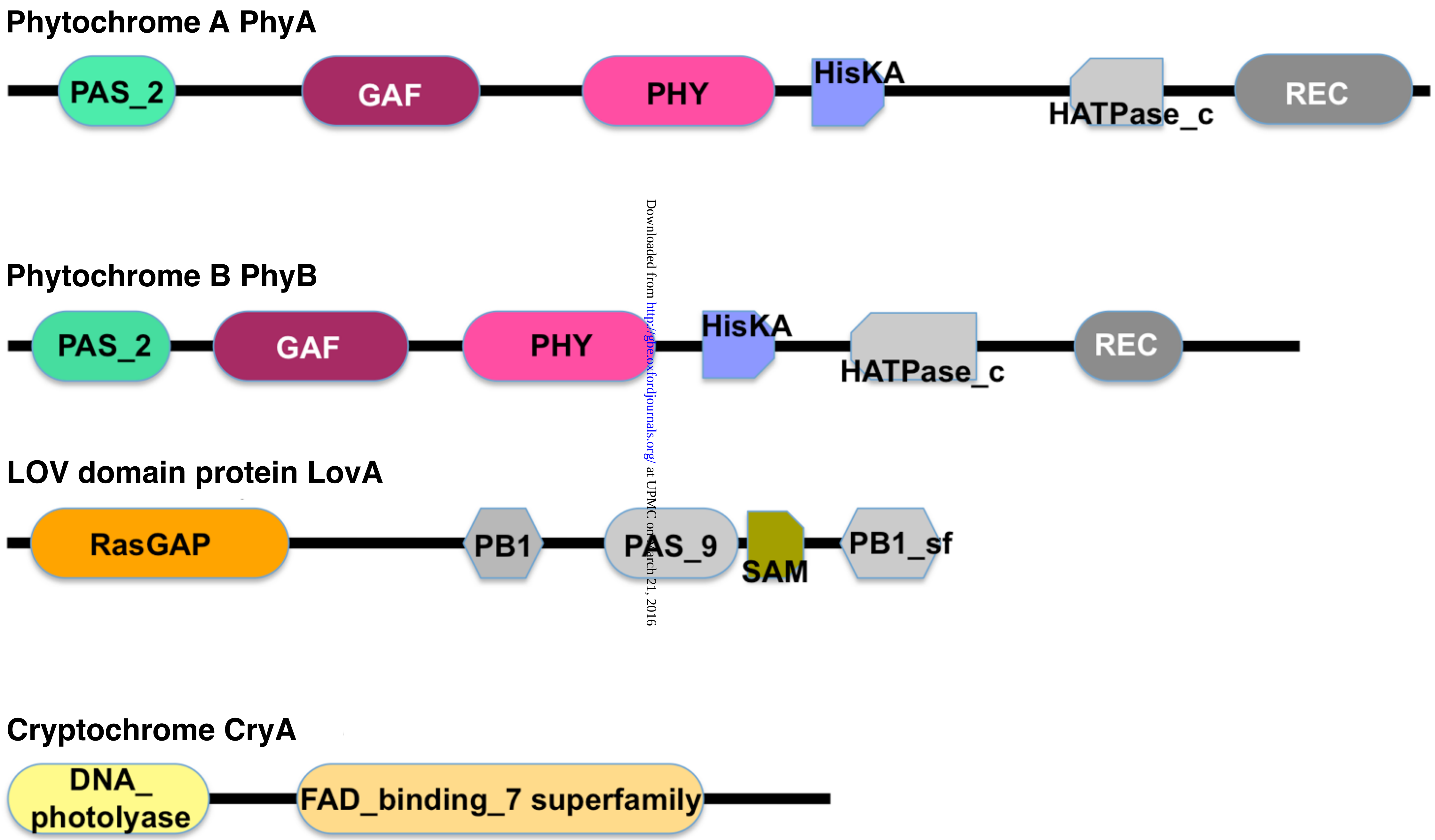
B. Receivers. Putative effectors for SHKP activated signalling were identified by query of genome and transcriptome sequences with receiver/response regulator protein seqeunces or the Interpro identifier IPR001789 for this domain, respectively, leaving out all proteins that also contained HisKA or HATPase-c domains. Query for outgroup sequences mainly retrieved SHKPs, which also contain a receiver domain, and *Dictyostelium* RegA. RegA and one other none SHKP outgroup sequence (YP_005167077) were aligned with the *P. polycephalum* sequences and a phylogenetic tree was constructed by Bayesian inference as described above. The tree was decorated with protein domain architectures.

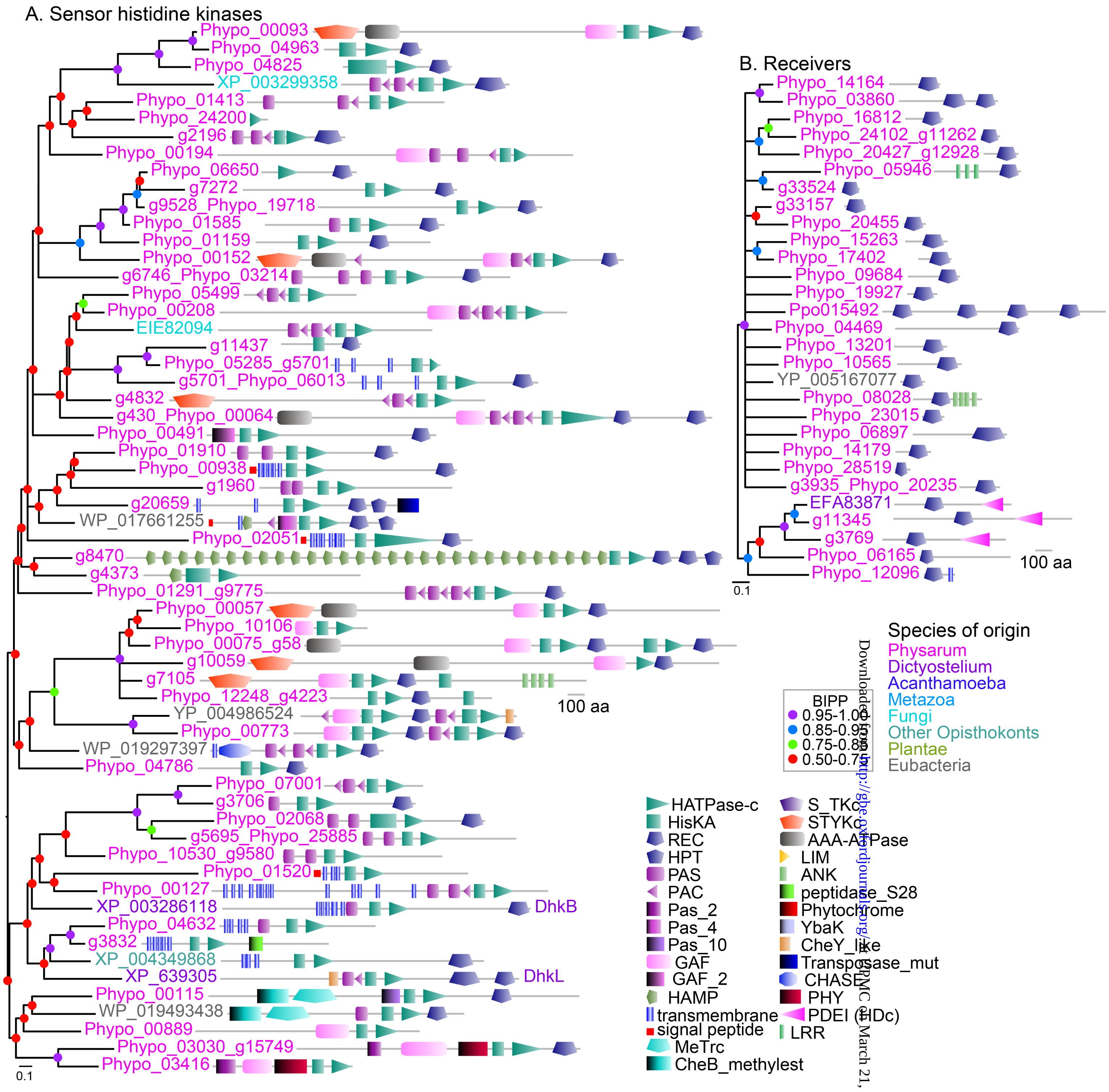
Figure 5. Alignment and phylogeny of putative tyrosine kinases. The sequences of the most consensual *P. polycephalum* tyrosine kinases were aligned with those of their closest homologs in other species. A) Section of the alignment that is essential for peptide substrate recognition, with residues essential for all protein kinases in blue text, for tyrosine kinases in red text and for S/T or S/T/Y kinases in green. The identifiers of *P. polycephalum* proteins with full tyrosine kinase consensus are underlined. B) Bayesian phylogeny constructed from the full kinase domain alignment, annotated with protein domain architectures. Species protein tags and node probabilities are colour coded as in Fig. 4. The methods described in the legend to Fig. S13 were used for phylogenetic inference. For *D. discoideum* homologs with confirmed substrate specificity, gene names and kinase designation (TK and TKL:tyrosine; S/TK serine/threonine) follow the protein tags.

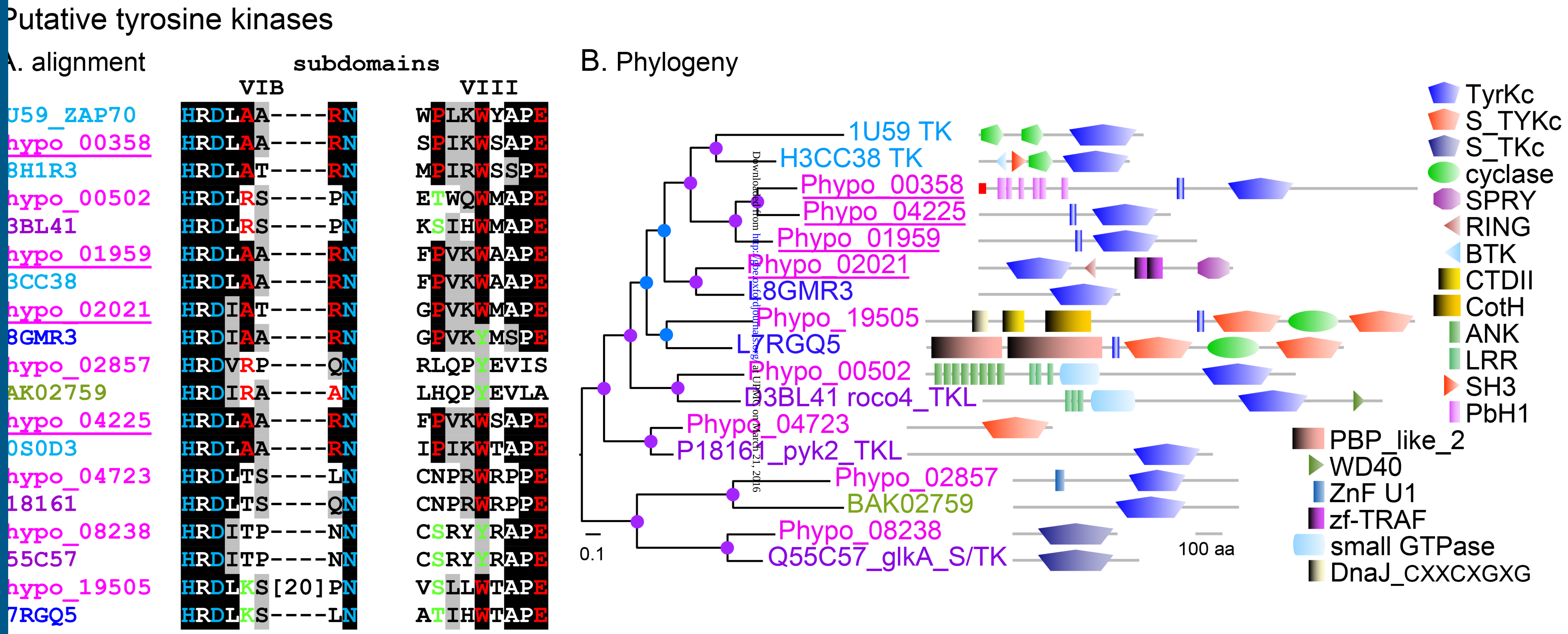
Figure 6. Proteins with SH2 domains. Proteins that harboured the Interpro IPR000980 SH2 domain were retrieved from an Interproscan of all transcribed coding sequences. The SH2 domain is too small for meaningful sequence alignment based phylogeny reconstructions and proteins are therefore classified by their protein domain architechture. The identifiers of the proteins of each type are as follows: A: Phypo_00702, Phypo_02696, Phypo_03176, Phypo_05058, Phypo_06094, Phypo_06144, Phypo_06676, Phypo_ 06719, Phypo_6732, Phypo_07245; B: Phypo_02646, Phypo_02863, Phypo_01571; C: Phypo_10425, Phypo_14943, Phypo_03389; D: Phypo_05516; E: Phypo_00177.

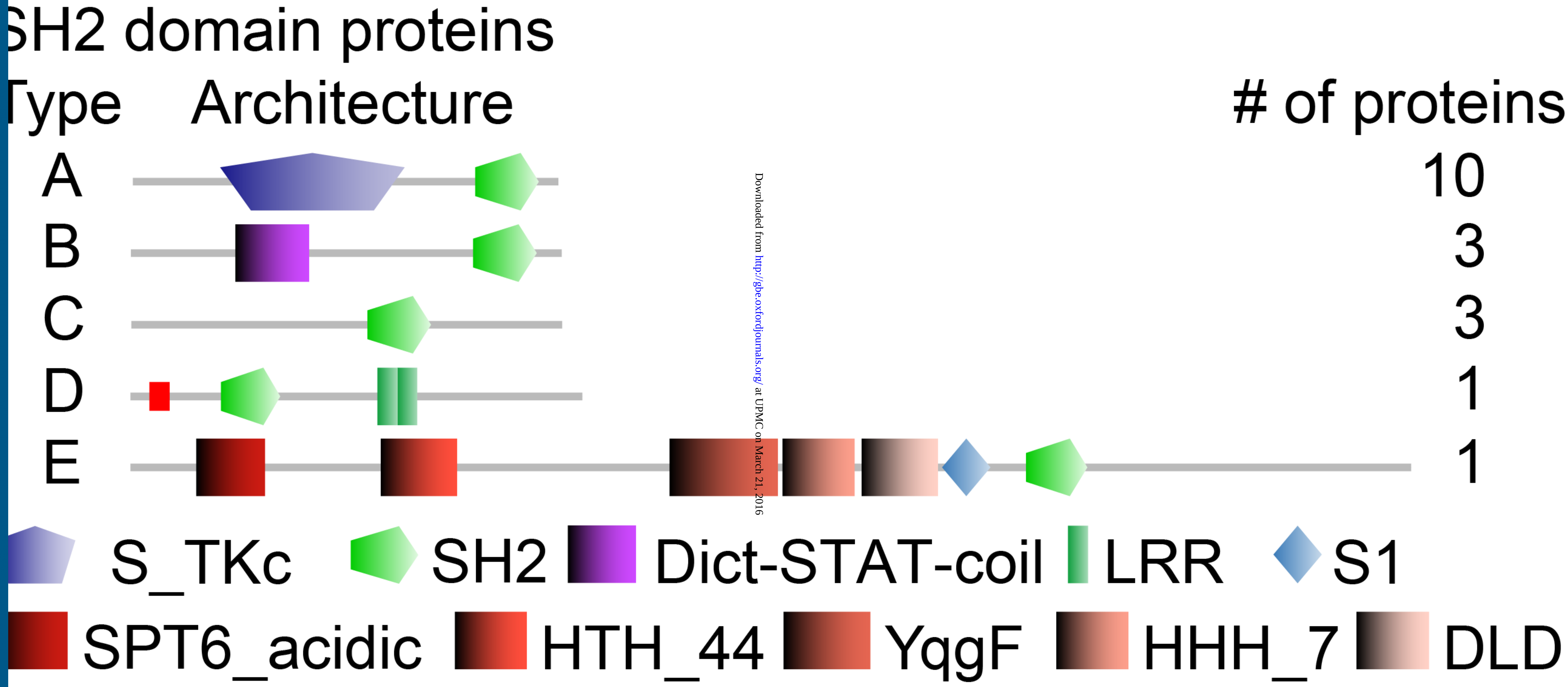

A
Sporangium
Meiosis
Spore
(n)
Amoeba
(n)
Flagellate
Sporulation
Apogamic cycle
Mating
Cyst
Apogamic
development
30 °C
Zygote
(2n)
(n)
Young
plasmodium
(2n)
Mature
plasmodium (2n)
Sclerotium
Heterothallic cycle
B
100
100
100
100
100
Dictyostelium discoideum
Dictyostelium purpureum
Dictyostelium lacteum
Dictyostelium fasciculatum
Polysphondilium pallidum
Physarum polycephalum
Acanthamoeba castellanii
Dictyostelia
Mycetozoa
Amoebozoa
0.05

			(B3/O)		(E/F/J/G/I)		(Myt1)		(Cdh1)				
	(H. sapiens)	**Cdk1-4,6**	**CycB**	**CycA**	**CycD**	**Cdc25**	**Wee1**	**Cdc20**	**Fzr1**	**E2F1-6**	**E2F7-8**	**DP**	**Rb**
Metazoa	*H. sapiens (human)*	5	4	2	12	3	3	2	1	6	2	3	3
	D. rerio (zebrafish)	5	4	2	11	2	2	1	2	5	2	3	3
	B. floridae (lancelet)	5	2	1	5	1	2	1	2	2	2	1	2
	D. melanogaster (fly)	3	2	1	3	2	2	1	2	2		1	2
	T. adhaerens (placozoa)	3	2	1	3	1	1	1	1	2	1	1	2
Holozoa	*Monosiga brevicollis*	1	1	1	2	2	1	1	1	1	1	1	1
Fungi	*Saccharomyces cerevisiae*	1	9			1	1	2	1				
	Candida albicans	1	5			1	1	2	1				
	Neurospora crassa	1	3			1	1	2	1				
	Schizosaccharomyces pombe	1	5			1	2	3	2				
	Ustilago maydis	1	3			1	1	1	1				
Apusozoa	*Thecamonas trahens*	1	1	1	3	2	1	1	1	1	1	1	1
Amoebozoa	*Dictyostelium discoideum*	1	1	1	1	1	3	1	1	1		1	1
	Dictyostelium purpureum	1	1	1	1	1	3	1	1	1		1	1
	Dictyostelium fasciculatum	1	1	1	1	1	3	1	1	1	1	1	1
	Polysphondylium pallidum	1	1	1		1	1	1	1	1	2		1
	Physarum polycephalum	**3**	**1**	**1**	**2**	**3**	**4**	**2**	**1**	**1**		**1**	**1**
	Acanthamoeba castellanii	1	2	1	2	2	2	5	1	1		1	1
Excavata	*Naegleria gruberi*	3	3	3		1	1	3	1	1	1	1	1
Cryptophyta	*Guillardia theta*	4	3	1	3	1	3	3		1	1	2	1
Haptophyta	*Emiliania huxleyi*	2	2	3	1		2	6	2	2			1
SAR	*Reticulomyxa filosa*	3	5	1		1	4	1	2				
	Bigelowiella natans	1	1	1	4	3	4	2	1	1	2		
	Phytophthora infestans	5	2	1	2	1	2	1	1	1	1	1	1
	Ectocarpus siliculosus	3	2	1	2		2	2	1	1	1	1	1
Archaeplastida	*Porphyridium cruentum*	2	1	1	1	1	3	1	1	1	1	1	
	Cyanidioschyzon merolae	2		1	1	1	1	1	1	1	1	2	1
	Ostreococcus tauri	2	1	1	2		1	1	1	1	1	1	1
	Coccomyxa subellipsoidea	4	1	1	1		2	1	1	1	1	1	1
	Selaginella moellendorffii	3	1	3	3		1	3	2	2	1	1	2
	Arabidopsis thaliana	5	11	10	11		1	6	3	3	3	2	1
	(A. thaliana)	**CdkA/B**	**CycB**	**CycA**	**CycD**		**Wee1**	**Cdc20**	**Fzr1**	**E2FA-C**	**DEL**	**DP**	**RBR**
					(SDS)								


Phytochrome A PhyA
PAS_2
GAF
PHY
HisKA
HATPase_c
REC
Phytochrome B PhyB
PAS_2
GAF
PHY
HisKA
HATPase_c
REC
LOV domain protein LovA
RasGAP
PB1
PAS_9
SAM
PB1_sf
Cryptochrome CryA
DNA_
photolyase
FAD_binding_7 superfamily


A. Sensor histidine kinases
Phypo_00093
Phypo_04963
Phypo_04825
XP_003299358
Phypo_01413
Phypo_24200
g2196
Phypo_00194
Phypo_06650
g7272
g9528_Phypo_19718
Phypo_01585
Phypo_01159
Phypo_00152
g6746_Phypo_03214
Phypo_05499
Phypo_00208
EIE82094
g11437
Phypo_05285_g5701
g5701_Phypo_06013
g4832
g430_Phypo_00064
Phypo_00491
Phypo_01910
Phypo_00938
g1960
g20659
WP_017661255
Phypo_02051
g8470
g4373
Phypo_01291_g9775
Phypo_00057
Phypo_10106
Phypo_00075_g58
g10059
g7105
Phypo_12248_g4223
YP_004986524
Phypo_00773
WP_019297397
Phypo_04786
Phypo_07001
g3706
Phypo_02068
g5695_Phypo_25885
Phypo_10530_g9580
Phypo_01520
Phypo_00127
XP_003286118
DhkB
Phypo_04632
g3832
XP_004349868
XP_639305
DhkL
Phypo_00115
WP_019493438
Phypo_00889
Phypo_03030_g15749
Phypo_03416
0.1
100 aa
B. Receivers
Phypo_14164
Phypo_03860
Phypo_16812
Phypo_24102_g11262
Phypo_20427_g12928
Phypo_05946
g33524
g33157
Phypo_20455
Phypo_15263
Phypo_17402
Phypo_09684
Phypo_19927
Ppo015492
Phypo_04469
Phypo_13201
Phypo_10565
YP_005167077
Phypo_08028
Phypo_23015
Phypo_06897
Phypo_14179
Phypo_28519
g3935_Phypo_20235
EFA83871
g11345
g3769
Phypo_06165
Phypo_12096
0.1
100 aa
Species of origin
Physarum
Dictyostelium
Acanthamoeba
Metazoa
Fungi
Other Opisthokonts
Plantae
Eubacteria
BIPP
● 0.95-1.00
● 0.85-0.95
● 0.75-0.85
● 0.50-0.75
HATPase-c
HisKA
REC
HPT
PAS
PAC
Pas_2
Pas_4
Pas_10
GAF
GAF_2
HAMP
transmembrane
signal peptide
MeTrc
CheB_methylest
S_TKc
STYKc
AAA-ATPase
LIM
ANK
peptidase_S28
Phytochrome
YbaK
CheY_like
Transposase_mut
CHASE
PHY
PDEI (HDc)
LRR


Putative tyrosine kinases
A. alignment
subdomains
VIB
VIII
U59_ZAP70 HRDLAA----RN WPLKWYAPE
hypo_00358 HRDLAA----RN SPIKWSAPE
8H1R3 HRDLAT----RN MPIRWSSPE
hypo_00502 HRDLRS----PN ETWQWMAPE
3BL41 HRDLRS----PN KSIHWMAPE
hypo_01959 HRDLAA----RN FPVKWAAPE
3CC38 HRDLAA----RN FPVKWAAPE
hypo_02021 HRDIAT----RN GPVKWMAPE
8GMR3 HRDIAA----RN GPVKYMSPE
hypo_02857 HRDVRP----QN RLQPYEVIS
AK02759 HRDIRA----AN LHQPYEVLA
hypo_04225 HRDLAA----RN FPVKWSAPE
0S0D3 HRDLAA----RN IPIKWTAPE
hypo_04723 HRDLTS----LN CNPRWRPPE
18161 HRDLTS----QN CNPRWRPPE
hypo_08238 HRDITP----NN CSRYYRAPE
55C57 HRDITP----NN CSRYYRAPE
hypo_19505 HRDLKS[20]PN VSLLWTAPE
7RGQ5 HRDLKS----LN ATIHWTAPE
B. Phylogeny
1U59 TK
H3CC38 TK
Phypo_00358
Phypo_04225
Phypo_01959
Phypo_02021
L8GMR3
Phypo_19505
L7RGQ5
Phypo_00502
D3BL41 roco4_TKL
Phypo_04723
P18161_pyk2_TKL
Phypo_02857
BAK02759
Phypo_08238
Q55C57_glkA_S/TK
0.1
100 aa
TyrKc
S_TYKc
S_TKc
cyclase
SPRY
RING
BTK
CTDII
CotH
ANK
LRR
SH3
PbH1
PBP_like_2
WD40
ZnF U1
zf-TRAF
small GTPase
DnaJ_CXXCXGXG


SH2 domain proteins
Type
Architecture
# of proteins
A
10
B
3
C
3
D
1
E
1
S_TKc
SH2
Dict-STAT-coil
LRR
S1
SPT6_acidic
HTH_44
YqgF
HHH_7
DLD