Deep origins of eukaryotic multicellularity revealed by the Acrasis kona genome and developmental transcriptomes

Sanea Sheikh
Uppsala University https://orcid.org/0000-0002-8615-2982
Chengjie Fu
Uppsala University
Matthew Brown
Mississippi State University
Sandra Baldauf
sandra.baldauf@ebc.uu.se

Uppsala University

Article

Keywords: development, signaling, multigene families, differential gene expression, autophagy, proteasome, kinetochore, flagella, horizontal gene transfer, extracellular matrix

Posted Date: March 1st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2587723/v1

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Communications on November 25th, 2024. See the published version at https://doi.org/10.1038/s41467-024-54029-z.

Abstract

Acrasids are large, fast-moving, omnivorous amoebae. However, under certain conditions, they can also cooperate to form multicellular fruiting bodies in a process known as aggregative multicellularity (AGM). This makes acrasids the only known example of multicellularity among the earliest branches of eukaryotes (formerly superkingdom Excavata) and thus the outgroup to all other known multicellular eukaryotes. We have sequenced the genome of *Acrasis kona*, along with transcriptomes from cells in pre-, mid- and post-development. We find the *A. kona* genome to be rich in novelty, genes acquired by horizontal transfer and, especially, multigene families. The latter include nearly half of the amoeba's protein coding capacity, and many of these families show differential expression among life cycle stages. Development in *A. kona* appears to be molecularly simple, requiring substantial upregulation of only 449 genes compared to 2762 in the only other AGM model, *Dictyostelium discoideum.* However, unlike the dictyostelid, developing *A. kona* also does not appear to be starving, being instead very metabolically active and inducing neither autophagy nor increasing ubiquitin-tagged proteolysis. Thus, contrary to current expectations, starvation does not appear to be essential for AGM development. Moreover, despite the ~ 2 billion years of evolution separating the two amoebae, their development appears to employ remarkably similar pathways for signaling, motility and construction of an extracellular matrix surrounding the developing cell mass. In addition, much of this similarity is shared with the clonal multicellularity of animals. This makes the acrasid something of a "bare bones" developmental model and suggests that much of the basic tool kit for multicellular development arose very early in eukaryotic evolution.

Introduction

Eukaryotes employ two basic modes of multicellularity. In clonal multicellularity, a single cell develops into a multicellular organism by coordinated growth and differentiation. In aggregative multicellularity (AGM), growth and differentiation occur separately, intersected by a striking transition from asocial growth to social development. Both strategies require cell-cell signaling, interaction and cooperation, and both have evolved multiple times. However, while clonal multicellularity is well studied in several very different systems, *e.g.*, plants and animals, AGM is only well studied in the dictyostelid amoebae, especially the developmental model *Dictyostelium discoideum.* We are studying AGM in the heterolobosean amoeba, *Acrasis kona*, which is separated from dictyostelids by over a billion years of evolution (Strassert et al. 2021).

Acrasids are large fast-moving amoebae, common in the wild and easily grown in the lab (Brown et al. 2010, movie_1). After several days on solid or liquid media, amoebae enter a social stage, with groups of cells aggregating in regions devoid of food. The resulting aggregate surrounds itself with an extracellular matrix and then commences to develop into a multicellular fruiting body (sorocarp, Fig. 1). Morphogenesis begins with the basal sorogen cells encysting individually to form a stalk composed of irregularly-shaped thick-walled cysts. The stalk continues to grow as more basal cells encyst, gradually lifting the remaining cell mass off the substrate. Once the stalk is complete, the aerial cell mass

proceeds to form lobes, which then elongate to form branching uniserrate rows of cells. With the aerial array in its final confirmation, the aerial cells proceed to encyst *en mass* to form uniformly rounded thick-walled spores. Encysted aerial cells (spores) are further differentiated from encysted stalk cells by the presence of raised, highly-pigmented, ring-like hyla at each spore-spore contact point (Olive 1975). Four *Acrasis* species have been described, each with a distinct sorocarp morphology, ranging from a simple stalk to the multiply branching tree-like structures of *A. kona* (Fig. 1; Brown et al. 2010).

The acrasid life cycle is strikingly similar to that of dictyostelids (Baldauf & Strassmann 2017), and acrasids were long considered their primitive relatives (van Tieghem 1880). Due at least in part to this presumed close relationship *Acrasis rosea* (now *Acrasis kona*) enjoyed some popularity as an experimental model, especially for the study of cytoskeletal and related features (e.g., Hellstén and Roos 1998). However, the amoebae of acrasids and dictyostelids differ markedly in both morphology and behavior (Olive 1975, Page and Blanton 1985), and molecular phylogeny now places them widely separated in the eukaryote tree. Thus, acrasids are now placed together with the model organism *Naegleria gruberi* in phylum Heterolobosea (suprakingdom Discoba; Pánek et al. 2017), while dictyostelids are placed with animals and fungi in suprakingdom Amorphea (Adl et al. 2019). We are studying the evolution of development using *A. kona* as a model system, beginning with the sequencing of its genome and three life-cycle stage-specific transcriptomes (Fig. 1). This allows us to model the critical transition from asocial feeding to social development in *A. kona*, and then to compare this to the corresponding transition in the dictyostelid model, *Dictyostelium discoideum* AX4 (Ddi AX4, Santhanam et al. 2015). Our results reveal a remarkable similarity in central developmental pathways shared by the acrasid, dictyostelid and animals.

Results

The Acrasis kona genome

Based on the genome assembly (Table S1), we estimate the *A. kona* genome to be 44.02 Mbp in size and essentially complete (92.7% CEGMA, 93.1% BUSCO; Table S2) (Parra et al. 2007, Simão et al. 2015). Roughly half of the *A. kona* genes are predicted to have introns, with an average of 1–2 introns per gene and a size range of 9–95 bp (Table S1), and 28% of the genes have predicted transmembrane domains (TMs) and/or signal peptides (SPs) (Table S3). The genome appears to be rich in novelty, with nearly a third of the 15868 predicted proteins having no GenBank BLASTp hits below e -5 (5987 accessions). Repeats are also numerous and diverse, covering 16.8% of the genome compared to 5.1% of the genome *Naegleria gruberi* (Table S4), the only other member of phylum Heterolobosea with a well-characterised genome (Fritz-Laylin et al. 2010). The *A. kona* predicted proteome is also highly redundant, with nearly half of the predicted proteins clustering into 2728 orthology groups (OGs) ranging in size from 2 to 49 members (Tables S5 and S6). Most of these families also seem to have evolved relatively recently as the vast majority are single copy or absent in *N. gruberi* (2369 families), and over half are single-copy or absent across a wide sampling of eukaryotes (1607 families, Table S6). Some of this redundancy may be related to the high content of membrane proteins, 88% of which are multicopy (Table S6). *A. kona* is also

rich in genes apparently acquired by horizontal gene transfer (HGT), and both multicopy and HGT genes are moderately enriched for metabolic functions (Figure S1, Table S7). The *A. kona* metabolic repertoire is diverse except for lacking anaerobic capacity, indicated by the absence of Fe-hydrogenase and pyruvate:ferredoxin oxidoreductase (Table S7, Figure S2).

Phylogenetic analysis of all *Acrasis kona* HGT candidates with widespread taxonomic distribution (multiHits) yields 1253 confirmed HGTs (cHGTs), with > 60% maximum likelihood ultrafast bootstrapping (UFB) support (Fig. 2A). Many HGT candidates found in only one non-excavate major taxon (uniHits, Fig. 2A) are probably also legitimate HGTs, particularly those that are widespread in the relevant non-excavate group. However, these HGTs cannot be confirmed by phylogeny due to the lack of an outgroup. For example, uniHits found exclusively in *A. kona* and diverse Bacteria (Fig. 2B) are almost certainly of bacterial origin, and the presence of introns (*e.g.*, Fig. 2C) and signal peptides in many of the *A. kona* accessions confirms that they are not bacterial contaminants (*e.g.*, Fig. 2D). *A. kona c*HGTs also tend to be redundant, with over half clustering into protein families (OGs). Some of these families are monophyletic indicating post-transfer expansion (196 accessions, Fig. 2A). However, the bulk of redundant *A. kona* cHGTs are in families of mixed origin (496 accessions, Fig. 2A), indicating either multiple transfers from different donors, or, most often, a combination of horizontal and vertical transmission (Table S8). Thus, while HGT is a notorious source of genetic novelty, much of the HGT in *A. kona* appears to contribute to redundancy. The acrasid cHGTs also trace to all domains of life but especially to clades dominated by soil microbes, specifically the Ciliophora, Plasmodiophora, Oomycota, and Fungi (Table S8), which are probably abundant on the dead or dying vegetation where acrasids are most commonly found (Olive 1975, Brown et. al 2010). This confirms that acrasids are omnivorous micro-predators (Olive 1975, Brown et. al 2010, movie_1), and they probably acquire HGTs mostly from their prey (Fig. 2A).

Intra- and extra-cellular signaling are critical for all organisms, including microbes. Social microbes have the added requirements of attracting aggregation partners, directing their movement in the aggregate and organizing them into a fruiting body. *Acrasis kona* encodes a wealth of signaling domains especially cyclase and calcium-binding (EF-hand) domains, RAS GTPases and hetero-trimeric G-protein regulators (Fig. 3; Table S9). The acrasid is also uniquely, if mildly enriched in nearly all components of phosphatidylinositol (PIP) signaling and blue light sensor (BLUF) domains, the latter also a powerful inducer of *A. kona* development (Olive 1975). Hybrid histidine kinases (hybrid hisK) stand out especially as far more abundant in *A. kona* than any of a diverse sampling of eukaryotes (Fig. 3). For example, *A. kona* has more than twice the number of hisK kinase (HK) and response regulator (RRR) domains (60 and 83, respectively) than its closest examined relative, *Naegleria gruberi* (28 and 33, respectively; Table S9). The acrasid also has a wealth of proteins with PAS domains, which are common hisKR interactors (IPR013767).

Developmental gene expression in Acrasis kona.

To gain insight into *A. kona* development, we sequenced transcriptomes from three key life cycle stages: growth (Gro), aggregation (Agg) and spore germination (Germ). These correspond to the pre-, mid- and

post-developmental states (Fig. 1). Gene expression was then compared between growth and aggregation (Gro vs Agg) and between aggregation and germination (Agg vs Germ), with substantial differential expression (SDE) defined based on a combination of length corrected read numbers (RPKM) and fold-change in expression (DE_{Log2}) (Table S10). Agg vs Germ corresponds to the initiation of growth and feeding (asocial trophic stage), while Gro vs Agg corresponds to the onset of development (non-feeding social stage). The latter comparison in *A. kona* appears to be roughly equivalent to five hours starvation in Ddi AX4 (Figure S3; Santhanam et al. 2015).

The bulk of differential expression among the life cycle stages examined here involves the transition to active growth, which shows increased (Germup) or decreased (Germdn) expression of 1771 and 2033 accessions, respectively (Fig. 4A). In contrast, the transition to development shows increased (Aggup) or decreased (Aggdn) expression of only 449 and 477 accessions, respectively (Fig. 4A). This contrasts rather dramatically with the 4337 accessions affected by aggregation in Ddi AX4 (Fig. 4A). Novelty among the *A. kona* accessions with increased expression during aggregation (Aggup) is also well below the genome average (~ 21% vs ~ 31%, Table 4A). Roughly half of the Germup and Germdn accessions are assigned to orthology groups (OGs), which is close to the genome average, while the percentage of Aggup and Aggdn accessions in protein families is again below average (43% vs 49%, Fig. 4A). Growth, in fact, appears to involve close to half of the *A. kona* protein families (1214 OGs, Fig. 4A). These families also tend to be specific to the acrasid, *i.e.*, less than a third are multicopy or even present in other eukaryotes (Fig. 4A, Table S6). Aggregation in contrast, uses more eukaryote-wide than unique protein families (88 out of 161 OGs, Fig. 4A). Thus, it appears that growth in *A. kona* is enriched for genetic redundancy and novelty compared to both development and the genome as a whole.

AGM is very widely, if sporadically, distributed across eukaryotes (Tice et al. 2016). One possible explanation for this could be horizontal transfer of some critical factor, most likely from the only diverse and ancient AGM taxon, the dictyostelids. Therefore, *Acrasis kona* Aggup accessions confirmed for horizontal transfer (cHGTs) are of particular interest (Table 4A). However, none of the 35 *A. kona* Aggup cHGTs trace to Dictyostelia based on BLASTp or phylogeny (Table S11). Moreover, these cHGTs tend to be involved in common housekeeping pathways, such as lipid and carbohydrate metabolism (Table S11). One possible exception is a phosphoinositide phospholipase C (Pi-PLCc) that appears to be of fungal origin (Figure S4). Pi-PLCc is a key enzyme in phosphoinositide (P_int) signaling, and evidence detailed below suggests that this pathway is involved in *A. kona* aggregation signaling.

Development in Acrasis kona versus Dictyostelium discoideum

The switch from solitary feeding to aggregation is considered the pivotal event in AGM development, and the Aggup repertoire is key to this shift. Nonetheless, only 449 accessions or less than 3% of the *A. kona* predicted proteome is Aggup. This contrasts sharply with Ddi AX4, where roughly a third of the proteome is substantially affected by aggregation and 2726 genes are Aggup (Table 4A). This suggests that

aggregation is molecularly much simpler in the acrasid than in the dictyostelid. To investigate this further, we annotated the *A.* kona Aggup accessions using a consensus of their top BLASTp hits in GenBank nr, human Swissprot, and Ddi AX4 RefSeq, along with linked data in the conserved domain database (www.ncbi.nlm.nih.gov/cdd/), InterPro, and DictyBase (dictybase.org; Fey et al. 2019). The accessions were then clustered into broad functional categories based on their consensus annotation, and their expression compared with that of their homologs in Ddi AX4, the only AGM organism for which developmental gene expression has been well-characterized (e.g., Santhanam et al. 2015, Glöckner et al. 2016). Comparisons were then made for active growth and feeding (0 h) versus advanced aggregation (5 h) (Fig. 4B).

The 449 *Acrasis kona* Aggup accessions show a broad functional profile, including large components of protein production (translation, protein modification/folding/sorting), carbohydrate chemistry and signaling, along with a large number of novel accessions (5 h, Fig. 4B). *A. kona* Aggup protein production includes 14 ribosomal proteins, which are some of the most highly expressed Aggup accessions, nine tRNA synthetases, five heat shock proteins and a nearly complete set of clathrin-coated vesicle subunits, some in multiple copies (Table S11). Aggregating *A. kona* also appears to have very active mitochondria, with Aggup accessions predicted to be involved in diverse mitochondrial functions including protein synthesis, the tricarboxylic acid cycle, ATP synthesis/transport and iron-sulfur cluster assembly (Table S11). Thus, in general, aggregating *A. kona* looks like a very active cell, despite the apparent lack of food. In fact, roughly half of the *A. kona* Aggup accessions, spread more or less evenly across the categories, are also expressed at RPKM > 10 during growth (0 h Fig. 4B). The strongest exception to this is novel accessions, less than half of which are detected at 0 h.

A very similar functional profile is seen for the DDI AX4 homologs of the *Acrasis kona* Aggup accessions during active growth, including similar proportions of accessions in nearly all functional categories (0 h; Fig. 4B). In fact, the two profiles are nearly mirror images of each other, with the exception of extracellular/secreted/defense proteins and, of course, novel proteins. However, after five hours without food, the Ddi AX4 profile diverges markedly (5 h, Fig. 4B). This includes massive reduction in most aspects of information processing, comprising most of the major components of translation such as nearly all ribosomal proteins and tRNA synthetases (Table S11). There is also a nearly complete loss of all aspects of mitochondrial maintenance and function, consistent with reports of mitochondrial autophagy in aggregating *Dictyostelium discoideum* (see below; Fischer & Eichinger 2019, Mazur et al. 2021). Thus, after five hours without food, developing Ddi AX4 looks like a starving organism, while developing *A. kona* does not. In fact, there are only three functional categories in which the two amoebae show similar and substantial numbers of homologous Aggup accessions - signaling, protein folding and sorting and cell shape and motility (cytoskeleton; Fig. 4B).

Starvation response, or not

AGM development in the lab begins when food is depleted or the amoebae migrate to food-depleted regions of the culture (Olive 1975). At this point, the cells are presumably beginning to starve. Most microbes respond to adverse conditions such as starvation by entering dormancy via encystation

(Schaap and Schilde 2018). However, AGM taxa can also respond by aggregating and building a multicellular fruiting body, potentially increasing the opportunity for, and extent of, dispersal. Eukaryotes use a number of pathways to gain energy and cellular building blocks when starved. These include autophagy, where cells digest their own proteins and/or organelles, and targeted proteolysis of ubiquitin-tagged proteins via the proteasome.

Autophagy is an ancient eukaryotic mechanism for cell survival during starvation (Kuo et al. 2018). It is also a major source of energy in aggregating Ddi AX4, particularly macroautophagy, the breakdown of mitochondria in vacuole-like autophagosomes (Fischer and Eichinger 2019). The universal components of autophagy are not well defined, as there is very low sequence conservation in many component proteins between the best studied model systems, yeast and human, and even some proteins with the same annotation show no discernible sequence homology (see below). Therefore, we first predicted a universal autophagosome based on annotation in four model organisms – *Homo sapiens*, *Saccharomyces cerevisiae S288C*, *Schizosaccharomyces pombe* and DdiAX4. The Ddi AX4 homologs were then used as BLASTp queries against *Acrasis kona* and diverse other eukaryotes (Fig. 5, Table S12).

Both Ddi AX4 and *Acrasis kona* appear to encode complete autophagosomes, with many component proteins showing high sequence conservation. This includes many accessions with strong BLASTp hits to taxa as distantly related as human ($< e^{-100}$, Fig. 5, Table S12). During growth, *A. kona* also shows similar expression patterns to Ddi AX4 for most of these proteins (Table S12). However, there is no discernible increase in expression, much less SDE, for nearly all of these components during aggregation in *A. kona* (Fig. 5). Moreover, autophagy overlaps with vacuolar assembly and transport, so the few *A. kona* SDE Aggup autophagy-like accessions could reflect other functions. In contrast, nearly all autophagy components are two- to four-fold Aggup in aggregating Ddi AX4 (Fig. 5, Table S12). Thus, *A. kona* does not appear to be using autophagy to generate energy during aggregation, while Ddi AX4 clearly does (Fischer & Eichinger 2019).

Another potential source of amino acid building blocks in starving cells is proteolysis, and the major cellular machinery for this is the proteasome (Bard et al. 2018). Again, *Acrasis kona* encodes a complete and highly sequence-conserved proteasome (Fig. 6, Table S13). Most of the component proteins are also moderately- to highly-expressed in growing and aggregating *A. kona* cells, especially the proteasomal regulatory subunit Rpn11 (1813 RPKM). However, only four of the 40 proteasomal proteins are Aggup in *A. kona*, while two are Aggdn (Fig. 6). In contrast, although Ddi AX4 shows moderate to moderately-high expression of proteasomal protein genes during growth, expression of almost every gene increases at least 2-fold during aggregation (Fig. 6, Table S13). Thus, Ddi AX4 markedly increases protein degradation during aggregation, while the acrasid does not. The *A. kona* proteasome does nonetheless appear to be tightly regulated, as the entire system appears essentially shut down in germinating spores (Fig. 6, Table S13).

One degradative system that appears to be very active in all three life-cycle stages of *Acrasis kona* is the exosomal RNA degradation machinery (Figure S5, Table S14; Kowalinski et al. 2016). This includes

Dis3L/RPR44, the RNAase component of the cytosolic exosome, for which there are over 4800 RPKM in all three stages (Figure S5). The cytosolic exosome plays central roles in translation regulation, mRNA quality control, transposon suppression, viral defense and gene regulation by RNA interference (RNAi) (Januszyk and Lima 2014). In fact, *A. kona* appears to encode nearly all universal components of RNAi (Łabno et al. 2016), some of which appear to be constitutively, if moderately, expressed in all life cycle stages examined here (Table S14). This may be related to the abundance of repeats in the acrasid genome (Table S3) and/or the fact that *A. kona* encodes 3–4 copies of the eukaryotic reverse transcriptase rvt, although the function of rvt is not known (Figure S6; Gladyshev & Arkhipova 2011). However, it appears unlikely that the exosome is a major source of nucleotide building blocks during starvation in either the acrasid or the dictyostelid.

Forty-three *Acrasis kona* Aggup accessions are annotated as cell-cycle related including accessions annotated for roles involved in regulation (4 accessions), cell division (8 accessions), DNA replication/repair (11 accessions) and nucleotide biosynthesis (10 accessions) (Table S15). This suggests a possible role for cell division in *A. kona* development. Cell division could be useful for taxa that struggle to find sufficient numbers of aggregation partners for a viable sorocarp (Romeralo et al. 2013). Therefore, we reconstructed the *A. kona* kinetochore and examined its expression (Figure S7, Table S15). *A. kona* encodes a largely sequence-conserved kinetochore (Tromer et al. 2019), although it appears that *A. kona* is unique among examined eukaryotes in lacking the key regulatory protein Cdh1 (Figure S8). Instead, the acrasid has two copies of the closely related Cdc20 (Figure S10), suggesting that the two proteins can functionally substitute for each other. However, there is little indication that aggregating *A. kona* is actively dividing. While a few kinetochore accessions show some increased expression during aggregation, a similar number show a decrease (6 Aggup, 7 Aggdn, Figure S8), and nearly all kinetochore accessions are expressed at only low to moderate levels (1.6–44 RPKM) (Table S16). There is also no evidence that the *A. kona* sorocarp contains more cells than the initial aggregate.

Nonetheless, *Acrasis kona* Aggup accessions include DNA polymerase subunits α and δ, five replication complex proteins (ginsA, DNA topoII, RNAse HII, mcm3, cdk45), and 13 cell-cycle related accessions including five regulatory kinases (cdk2, cdk5, cyc_U4, Aurora, nimA). Many of these accessions are also moderately to highly expressed and/or Germdn (Table S15). Thus, although DNA replication is probably not leading to mitosis in aggregating *A. kona*, reaching an advanced cell-cycle stage may still be important. Cell-cycle stage influences cell fate in Ddi AX4, with cell-cycle advanced cells more likely to end up as viable spores as opposed to dead stalk cells. This is apparently because late cell-cycle cells make more robust spores (Chen and Kuspa 2005). The possibility for a related phenomenon in *A. kona* is suggested by its large Aggup component of DNA replication and cell cycle accessions. Although all aggregating *A. kona* amoebae form viable spores or cysts (Olive 1975), the two cell types may have quite different fates. For example, aerial spores are more likely to be dispersed and to do so individually, while stalk cysts are probably more likely to remain behind as a group. This could lead to a selection process related to the suitability of different cell-stage cells for different fates.

Developmental Signaling

Cell signaling, both within and between cells is critical for development, and a broad sampling of the *Acrasis kona* extensive repertoire of signaling domains (Fig. 3) is Aggup (Fig. 7, Table S17). The *A. kona* Aggup external signaling repertoire includes a single G-protein coupled receptor (GPCR) and two G-protein beta subunits (Gβ), one complete and one partial hybrid histidine kinase receptors (hybrid hisKR), two blue light receptor proteins (BLUF), two plasma membrane (PM) calcium pumps (Ca-ATPase_IIb), two lipocalin-interacting receptors (LMMBR) and three predicted single transmembrane domain (1TM) receptors (Fig. 7). Of these, the GPCR, both hisKRs and two of the 1TM receptors appear to be largely aggregation specific (Germdn – 0.98 to -2.81, Table S17). *A. kona* is particularly rich in histdine kinase signal receptors (Fig. 3), although the two Aggup hisKRs are unusual in showing strongest similarity to bacterial sequences (e-value 0.0, Table S17). The BLUF proteins are likely involved in diurnal timing, and their increased presence during aggregation is consistent with the fact that blue light induces *A. kona* development (Reinhardt 1968).

Internally, aggregating *Acrasis kona* upregulates major components of the cyclic AMP (cAMP), phosphoinositide (P_int), sphingosine (S1P), mitogen activated protein kinase (MAPK), target of rapamycin (TOR) and assorted small GTPase pathways, albeit with some unique variation (Fig. 7). *A. kona* cAMP signaling most likely uses an essentially constitutive membrane-bound and sequence-divergent adenylate cyclase (AK_08377) along with Aggup versions of the main cAMP interacting partners CAP (adenylate cyclase associated protein), PKA (cAMP-dependent kinase) and two cAMP degrading phosphatases (PDE). For the P_int pathway, *A. kona* upregulates two versions of the key enzyme phospholipase C (PLC), which generates the second messengers inositol 3-phosphate (IP_3) and diacylglycerol (DAG). One of the main functions of IP_3 is binding to the IP_3 receptor in the endoplasmic reticulum (ER) membrane. This results in a flush of calcium release from the ER, thought to be the major site of calcium storage in eukaryotic cells. Although, there is no apparent ER IP_3 receptor among the *A. kona* Aggup proteins, nor was any detected by BLASTp search of the full *A. kona* proteome, the protein is not particularly sequence-conserved, and there are numerous novel multi-TM predicted proteins in the *A. kona* genome that could possibly fill this role (Table S3).

These signaling pathways appear to be tightly regulated as their Aggup components include both synthetase/activators and degradase/de-activators, specifically the cAMP, S1P, Ras, and TOR (Arf) pathways. Several of these pathways also converge on similar targets. Both the cAMP and TOR pathways likely affect cytoskeletal organization, suggesting changes in cell motility during *Acrasis kona* aggregation. This is despite the fact that no marked outward changes in motility are observed such as elongation toward the direction of aggregation as seen in Ddi AX4. Meanwhile, both the RAS and S1P converge on the MAPK signaling cascade, including two Aggup MAPK homologs. The most likely targets of these kinases are regulators of gene expression (Kranenburg and Moolenaar 2001).

External aggregation signaling is expected to be largely taxon-specific as it involves attracting and screening genetically related cells. Consistent with this, the dictyostelid and putative acrasid Aggup external signal receptors are more analogous than homologous (Fig. 7). Internal developmental signaling, on the other hand, shows extensive homology between the two amoebae including major

components of P_int, cAMP, SP1, RAS, TOR and MAPK signaling (Fig. 7, Table S17). Both amoebae also induce both internal and plasma membrane calcium pumps suggesting the possibility that both amoebae may use calcium as a chemoattractant (acrasin). Calcium is in fact a potent attractant for Ddi AX2, nearly as effective as their better-known acrasin, cAMP, albeit over shorter distances and time scales (Scherer et al. 2010). Calcium signaling alone is probably insufficient to assemble the large aggregates of *Dictyostelium spp*. However, it could be sufficient for the smaller acrasid aggregates, which also tend to be found on physically limited substrates (Olive 1975, Brown et al. 2010). It is also interesting to note that, although all species of the genus *Dictyostelium* appear to use cAMP as an acrasin, the bulk of dictyostelid diversity, most of which build much smaller sorocarps, do not aggregate in response to cAMP, and their actual acrasins are largely unknown (Romeralo et al. 2013).

The extracellular matrix

Extracellular matrices (ECM) are central to membrane-level interactions in both clonal and AGM development (Siu et al. 2004). In acrasids and dictyostelids, the sorogen and sorocarp are enclosed in an ECM referred to as the slime sheath. The dictyostelid ECM appears to be something of a cellular dumping ground consisting of a wide assortment of proteins and protein fragments in a carbohydrate matrix. These proteins include enzymes with obvious ECM functions such as carbohydrate metabolism and assorted proteases. However, in addition, there is a veritable alphabet soup of ~ 350 proteins or protein fragments with no obvious extracellular function, such as numerous ribosomal proteins (Bakthavatsalam and Gomer 2010). This has led to the suggestion that most of these proteins serve mainly to “bulk out” the ECM (Huber and O’Day 2015, Bakthavatsalam and Gomer 2010).
Acrasis kona Aggup accessions include sixty predicted proteins that are potentially involved in ECM construction. These include 23 accessions that together encode nearly all major components of COP, clathrin coated vesicle and vacuolar protein sorting/secretion systems, including multiple copies of several components of each (Table S18). There are also six potentially secreted Aggup peptidases, two extracellular matrix stabilizing proteins and at least six probably-secreted Aggup proteins with ankyrin or TPR repeats that could serve as protease targets. The *A. kona* Aggup peptidases include a highly-expressed aggregation-specific C26 peptidase, a predicted outer PM anchored zinc peptidase, an M60 peptidase, a tagC homolog and two strongly expressed potentially secreted M49 di-peptidases (Table S19). Thus, around 13% of the aggregating cell protein production is potentially involved in ECM construction. TagC is also interesting because of its critical role in the dictyostelid ECM, where it cleaves the pre-peptide SDF to release two small peptide hormones that control cell fate (Wang et al. 1999, Loomis 2014). Overall, then, the acrasid SDE Aggup accessions could include a full suite of proteins to construct and maintain an ECM and in a manner similar to that of dictyostelids and other multicellular eukaryotes (Huber and O’Day 2015).

Multigene families

The *Acrasis kona* genome is rich in multigene families (ortholog groups or OGs), with 7766 accessions grouped into 2728 OGs (Fig. 4A, Table S6). The bulk of these families are absent or single copy in other Discoba, and many form molecularly shallow clades (*e.g.*, Figures S4 and S6). This suggests that gene

family expansion has been very active in the acrasid lineage although due to an unusual combination of gene duplication and HGT (Fig. 2A). Well over a third of these families (1124 OGs) also show differential expression among the life cycle transitions examined here (Table S19). Accessions belonging to multigene families make up close to half of the SDE accessions for the Gro v Agg and, especially, the Agg v Germ transitions (42.9 and 50.4%, respectively, Fig. 4A). For Aggup, the bulk of the OG accessions are the only Aggup member of their family (Table S20), implying that other family members are expressed at different stages or under different growth conditions. However, twenty multigene families have multiple SDE Aggup accessions, and fourteen of these are two member families for which both members are Aggup (Table S20). More extreme examples are seen in the Agg v Germ transition. For example, OG3059 is unique to *A. kona* and consists of 23 hypothetical 6-transmembrane domain proteins of which nine are SDE Germup, and OG7439 for which all seven members are SDE Germup (Table S20).

Discussion

We have sequenced the genome of the aggregating amoeba, *Acrasis kona*, along with transcriptomes from the three main life cycle stages (Fig. 1). The ~ 44 Mb genome is predicted to encode 15868 proteins, ~ 1/3 of which are novel (Table 1) and nearly half of which cluster into multigene families. The amoeba has broad metabolic (Figure S2) and signaling (Fig. 3) repertoires, as well as a large and functionally diverse set of cHGT genes of diverse origins (Fig. 2). In the switch from active feeding to aggregation, *A. kona* shows substantially increased expression of less than 3% of its genome and little change in metabolic activity (Fig. 4). This is in sharp contrast to the dictyostelid model Ddi AX4, where aggregation affects nearly a third of the genome (Fig. 4A), and the cells appear to be starving (Figs. 5 and 6). Nonetheless, the two amoebae show very similar signal transduction profiles during aggregation (Fig. 7) and may use very similar enzymes and pathways to construct an extracellular matrix (Table S18).

Acrasis kona development

The 449 *A. kona* Aggup accessions can be roughly classified into three groups: novel genes, genes annotated for developmental pathways and housekeeping genes (Table S11). The main annotated developmental pathways are cell signaling, extracellular matrix construction and motility, and these are the main areas of homology between *A. kona* and Ddi AX4 Aggup genes. Moreover, homologs of many of these also play central roles in metazoan development (Pires-daSilva and Sommer 2003). The strong conservation of these sequences from acrasids to dictyostelids and human indicates that they are very old and probably play essential roles in many eukaryotes, including strictly microbial taxa. This is further supported by the fact that about half of these “developmental accessions” are not aggregation specific in the acrasid, *i.e.*, their expression is not substantially decreased in germinating cells (Germdn <-0.9) or even increases (Table S11). Thus, the acrasid appears to use ancient universal eukaryotic pathways for development, with stage-specific roles induced by a few key aggregation-specific components (Fig. 7).

This points to the likely importance for acrasid development of the small set of novel Aggup accessions, particularly the 36 that are aggregation specific (Table S21). Of particular interested are four predicted membrane receptors, three external membrane-anchored proteins, and three possibly secreted proteins. Eight of these accessions are also abundantly expressed (RPKM > 100), and 16 belong to multigene families, including multiple representatives of four families (Table S21). Altogether, these suggest that a minimum amount of invention was needed for the evolution of acrasid multicellularity. Expanded multigene families could also be considered a form of novelty. Gene duplication can facilitate evolutionary experimentation including developmental stage-specific timing and functional specialization, which also played major roles in the evolution of development in both plants and animals (Nei and Rooney 2005, Fernández and Gabaldón 2020). *A. kona* is rich in multigene families, many of which may have evolved relatively recently as they consist of genes that are either absent or single copy in a taxonomic diversity of eukaryotes (Fig. 4A, Table 11). Although *A. kona* Aggup accessions are not enriched for multicopy genes relative to the other life cycle stages examined here (Fig. 4A), over 190 Aggup accessions belong to multigene families. Moreover, for the vast majority of these accessions, they are the only Aggup family member (Table S21). Thus, novelty and differential expression of multigene families were probably central to the evolution of multicellularity in acrasids.

Acrasis vs Dictyostelium

The developmental cycle of acrasids and dictyostelids are remarkably similar. Both amoebae are induced to develop in the lab by starvation, at which point growth ceases and aggregation begins. During aggregation the cells gather together to build a mound of ECM-encased cells (the sorogen), which either continues to migrate as a unit or develops *in situ*. Once stationary, the sorogen transforms into a sorocarp consisting of a stalk supporting an aerial spore mass. In acrasids and the majority of dictyostelids, the sorocarp stalk is cellular, and aerial and stalk cells are differentiated, although dictyostelid stalk cells are dead at maturity while acrasid stalk cells are not. Cell fate is also not random in dictyostelids, with more cell-cycle advanced cells being more likely to survive to form spores (Chen & Kuspa 2005).

The possibility of a similar selective mechanism in *Acrasis kona* is suggested by the relatively large number of Aggup cell-cycle related accessions (34 accessions, Table S11), and the fact that the morphological differences between mature acrasid stalk and aerial cells likely affects their eventual fate. Mature acrasid stalk cells (stalk cysts) are irregularly shaped and sometimes expanded basally or parenchymatous (Olive 1975), while mature aerial cells (aerial spores) are smooth and rounded. Moreover, aerial spores are only joined to one another by single small hilar structures between pairs of adjoining spores, which may facilitate disintegration of the mature aerial array. This likely results in different fates for aerial spores versus stalk cysts, with the spores more likely to be dispersed but to do so individually, while the cysts most likely remain behind and as a unit. This would make stalk cysts more likely to survive and with sufficient neighbors to readily form a future sorocarp. Thus, there is the potential for a selection process contributing to cell-fate in the acrasid sorocarp.

However, there are also major differences between acrasid and dictyostelid development. Acrasid aggregation involves less than a hundred cells that migrate individually to form the sorogen, while dictyostelid aggregation is generally a highly coordinated affair with 100s to 1000s of cells gathering into distinct streams that gradually coalesce in predictable patterns (Raper 1985, Kawabe et al. 2019). Dictyostelid sorogens can then migrate considerable distances, while sorogen migration in acrasids is very limited, when it even occurs (Olive 1975). Dictyostelid spores also adhere to each other and disperse as a unit, so that a new locale is founded by a colony of closely-related cells rather than individual spores. The acrasid amoebae themselves are aggressive micro-predators, much larger (~ 8–10 fold) and faster than dictyostelids, and probably mostly prey on eukaryotic microbes (Fig. 2A, Olive 1975, movie_1), while dictyostelids are primarily bacteriovores (Raper 1985).

We also find that gene expression during development is markedly different between the acrasid and dictyostelid. After five hours without food, Ddi AX4 has substantially increased or decreased expression of nearly a third of its coding capacity (Eichinger et al. 2005), compared to less than 6% for *Acrasis kona* (Fig. 4A). Moreover, much of the Aggup response in *A. kona* concerns information processing and housekeeping proteins, nearly all of which are strongly downregulated in aggregating Ddi AX4 (Fig. 4B). Aggregating *A. kona* also does not induce autophagy or increase ubiquitin-linked proteolysis, while aggregating Ddi AX4 does both (Figs. 5 and 6). In general, then, aggregating *A. kona* looks like a very active cell undergoing a major life cycle transition with large increases in protein and energy production, while aggregating Ddi AX4 appears to be in survival mode. This suggests that much of the difference in magnitude between the two amoebae's aggregation response may be attributable to a complex starvation response in the dictyostelid. This probably includes additional pathways besides autodigestion, such as switching to more energy efficient forms of information processing and other housekeeping functions. In fact, the number of *A. kona* Aggup genes we report here (449 accessions, Fig. 4A) is remarkably similar to the number of developmentally essential genes estimated for Ddi AX3 when development is induced in the presence of food using rapamycin (~ 500 genes, Jaiswal and Kimmel 2019). Thus, while dictyostelid development is at least moderately more physically complex than that of *A. kona*, much of this may come down to a fairly small number of genes.

External developmental signaling

Extracellular signaling is central to AGM. For *Dictyostelium discoideum*, and probably all other known dictyostelids, this includes attracting cells, assessing co-aggregants for relatedness, monitoring aggregation size and organizing the cells into a sorocarp. For attracting co-aggregants, Dictyostelids use small diffusible molecules (acrasins), which is cAMP in the case of *Dictyostelium spp.* However, for most of the diversity of Dictyostelia, the acrasin appears to be glorin, folate or, in most cases, an unknown or undetermined molecule (Romeralo et al. 2013). However, calcium appears to be almost as powerful an acrasin as cAMP for Ddi AX2, at least over short distances and time scales (Scherer et al. 2010). All that is currently known of the *Acrasis kona* acrasin is that it is not cAMP (Olive 1975). One possibility suggested by the results here is that this role could be played by calcium. *A. kona* Aggup accessions include two versions of the IP_3 synthesizer PLC, one of which is highly aggregation specific (AK_13406),

and two strongly expressed plasma membrane calcium pumps (Fig. 7). This could be sufficient for the release of a pulse of calcium ion from the cell into the extracellular space. Moreover, the Ddi AX4 homologs of these enzymes are also Aggup (Fig. 7). Calcium cannot be relayed and amplified as efficiently as a degradable acrasin such as cAMP, which is alternately synthesized and degraded to create outwardly radiating waves of signal in the dictyostelid (Kriebel and Parent 2004). However, calcium signaling might be sufficient for gathering the relatively small numbers of cells needed to build an acrasid sorocarp, perhaps further assisted by the often physically limited substrates where acrasids are most commonly found (Olive 1975, Brown et al. 2010).

Dictyostelids screen their co-aggregants for relatedness using antigen-like cell-surface “TIGR” proteins, with the result that only closely related cells cooperate to form a sorocarp (Strassmann et al. 2011). This is should be especially important in dictyostelids, where roughly 20% of the cells are sacrificed to form the dead cellulosic stalk, making it disadvantageous for cells to risk being sacrificed for distant relatives. Although no cells are sacrificed to build the acrasid sorocarp, it is likely that the cells still perform some type of screening of co-aggregants for relatedness or at least the right species. This might be especially important if it uses Ca^{+} to attract aggregation partners, since the strong sequence conservation of its component proteins suggests that it is an ancient pathway and possibly universal among eukaryotes (Fig. 7). *A. kona* Aggup accessions include three novel proteins predicted to be anchored to the external cell surface by a single transmembrane domain (AK_09152, AK_00315, AK_10358), and all three are aggregation specific (Aggup 1.33–2.39. Germdn − 1.24 - -2.03). AK_09152 is also modestly highly expressed at 122 RPKM (Table S20 and S21). Although the latter still seems to be a fairly low level for a protein destined to be distributed across the surface of a large cell, cell-cell recognition is probably most important early in aggregation so that the responsible gene would be expressed at an earlier stage than examined here.

For monitoring aggregation size, *Dictyostelium discoideum* uses two different systems. During early development, amoebae use the presence in the medium of a glycoprotein secreted by starving cells (conditioned medium factor) to assess the density of nearby aggregation-ready cells (Gomer et al 2011). Later in development, the dicyostelid uses a 450 kD counting factor complex to assess aggregate size, leading to the splitting of overly large aggregates to form separate fruiting bodies (Gomer et al. 2011). *Acrasis kona* probably also assesses cell density, both early and late in aggregation. Before aggregation begins, acrasids can simply encyst individually if conditions are adverse but cell density is low (Olive 1975). Later in development, the acrasid may split an overly large aggregate into fragments that migrate apart to form separate sorocarps, with the result that *A. kona* sorocarps tend to be similar in size even in dense cultures (Olive 1975). The most obvious candidates for possible quorum sensors in *A. kona* are a hybrid hisK receptor/kinase and a second separate hisK kinase, both of which are over 4-fold Aggup and 2-fold Germdn (Fig. 7, Kleerebezem et al. 1997). Other possible candidates are three novel accessions (AK_11986, AK_08976, AK_14437), all of which are predicted to encode single transmembrane domain (1TM) receptors and have largely aggregation-specific expression (Aggup 1.14–2.39, Germdn − 1.19–2.81, Table S20). AK_11986, in particular, is moderately strongly expressed (331 RPKM) and predicted to form large internal and external membrane tails (Figure S9).

In fact, quorum sensing could be the main inducer of development in *Acrasis kona*, as the aggregating cells in the lab, at least, are probably not starving. *A. kona* commonly aggregates in the presence of food simply by cells migrating to the edges of the culture. This has the dual effect of bringing cells together, and probably doing so under more controlled conditions than in the midst of actively feeding cells and without the interference of live prey. Thus, the induction of aggregative multicellularity, at least in the case of *A. kona*, may be largely a matter of having sufficient cells to form a viable sorocarp. There is no obvious reason why this could not also occur in dictyostelids. Although starvation is a strong inducer of AGM in the lab for all known species of Dictyostelia (Kawabe et al. 2019, Fischer and Eichinger 2019), whether it is the only inducer in the wild seems unlikely. Spore formation and dispersal are probably useful means of survival under a variety of adverse conditions, in which case having sufficient cells to form a sorocarp could be the prime limiting factor.

Internal developmental signaling

While extracellular aggregation signaling is largely analogous between *Acrasis kona* and Ddi AX4, *A. kona* internal aggregation signaling appears to be largely homologous between the two amoebae. Like Ddi AX4, *A. kona* appears to use the P_int, MAPK, cAMP, Ras, TOR and S1P signaling pathways during aggregation, and most of the major components of these pathways are substantially up-regulated or constitutively expressed in both amoebae (Fig. 7, Table S18). Many of these pathway components also show strong sequence conservation despite the large evolutionary distance between Discoba and Amorphea. Thus, *A.* kona Aggup sequences for at least one key enzyme of the cAMP, TOR and S1P pathways match human and dictyostelid homologs with BLASTp e-values of < e-100, and both *A. kona* Aggup MAPKs hit both taxa with < e-75. In fact, these pathways also play central roles in animal development (Pires-daSilva and Sommer 2003, Patel and Shvartsman 2018). Thus, these appear to be very old among eukaryotic pathways, pre-dating the last common ancestor of actively-mitochondriate eukaryotes (Al Jewari & Baldauf 2022), at least, after which they have been repeatedly co-opted for roles in multicellular development.

The universality and strong sequence conservation of these signaling pathways indicates that they must play essential roles in eukaryotic processes unrelated to development. In fact, in *Acrasis kona* only a few key components of these pathways appear to be strongly aggregation specific, while most of the components are expressed at the same or even higher levels during germination (Fig. 7, Table S17). Nonetheless, at least one component of each pathway does appear to be aggregation specific (Aggup > 0.9, Germdn <-0.9). Thus, the acrasid appears to be regulating essentially constitutive signaling pathways using life cycle stage-specific induction of a few key non-constitutive components. Nearly all of the *A. kona* predicted membrane receptors also appear to have strongly life-cycle stage specific expression with the single exception of the blue light receptors (Fig. 7).

Evolution of development in Acrasis kona

Aggregative and clonal multicellularity together have evolved over twenty times in eukaryotes (Sebé-Pedrós et al. 2013), including at least once in nearly every major taxonomic division (Table S22). The

eight known instances of AGM are also likely an underestimate as sorocarps are mostly too small to detect in the field, and many species may not readily aggregate in the lab. Most known instances of AGM also involve, at most, a small group of closely related species with close non-aggregating sister clades. This suggests that AGM may be relatively easily evolved and largely ephemeral. This makes Dictyostelia, with an estimated age of 600 million years (Fiz-Palacios et al. 2013), over 150 described species, and probably a large hidden diversity (Baldauf et al. 2018), the most evolutionarily successful AGM clade by far. Whether this is because of or despite its developmental starvation response is an intriguing question. Meanwhile, although clonal multicellularity has given rise to at least five large diverse and ancient clades - animals, fungi, red algae, land plants, and brown algae (Sebé-Pedrós et al. 2013), there are also many examples of small experiments with clonal multicellularity, *e.g.*, among the volvocalean algae (Prochnik et al. 2010).

Acrasids are separated from all known multicellular taxa, including dictyostelids, by 1–2 billion or more years of evolution (Strassert et al. 2021). Nonetheless, the results here suggest that there is extensive similarity at the molecular level between development in both these taxa. In fact, both animals and fungi have close AGM relatives, *Capsaspora owckzarzaki* in the case of animals and *Fonticula spp.* in the case of fungi (Sebé-Pedrós et al. 2013). The similarities in fundamental developmental processes among these taxa suggest that much of the basic tool kit for multicellularity evolved early in eukaryotes. This includes internal and external signaling pathways, ECM construction, and probably also differential expression of multigene families. Thus, the taxonomically-wide, if sporadic distribution of multicellularity across eukaryotes suggests that they have been experimenting with multicellularity for much of their history. In which case multicellularity is either surprisingly rare, or there are many examples yet to be discovered.

Whatever key factors led to the evolution of acrasid multicellularity, the result is that the molecular biology of its development is not overlaid by a complex starvation response. Moreover, the number of genes involved appears to be small, a large fraction of which are performing simple housekeeping functions. Thus, the evolution of multicellularity in *A. kona* may have required little more than a few unique inventions against a large background of genomic redundancy. This makes *Acrasis kona* a very simple model for the study of universal eukaryotic pathways and how to they have been repeatedly co-opted for the evolution of multicellularity.

Experimental Procedures

Cell Culture and DNA Extraction

Acrasis kona ATCC strain MYA-3509 (Brown et al. 2012a) was grown on *Saccharomyces cerevisiae* in liquid culture or on CM+ (Corn Meal Plus) agar plates. For DNA extraction, spores from mature *A. kona* sorocarps were inoculated and grown in Spiegel's liquid medium (Spiegel 1982) in 250 ml flasks shaken at room temperature on a rotary shaker (120 cycles/min). Acrasid cells were harvested in 50 ml corning tubes after 48 h at a cell density of approximately 1×10^5/ml, at which point the yeast cells had

flocculated. Harvested cells were transferred to Petri dishes without food and left for at least 1 hour to allow the amoebae to complete digestion of residual yeast material and to settle and attach to the bottom of the plate. Cells were then washed three times with 10 mM phosphate buffer to remove the pellets of flocculated yeast and harvested by centrifugation. DNA was extracted using the Blood & Cell Culture DNA Kit (Qiagen) as described in Fu et al. (2014).

Genome sequencing and assembly

Sequences were generated from *A. kona* total DNA using 454 GS Titanium (Roche) and Genome Analyzer (Illumina) platforms. Library generation and sequencing for both the 454/Roche (FLX + shotgun) and Illumina systems (500 bp Pair-End & 2 kb Mate-Pair) were carried out according to the manufacturers' protocols. All reads were treated for base correction and quality trimmed using the FASTX toolkit (v0.0.13) (hannonlab.cshl.edu/fastx_toolkit/) and Trimmomatic (v0.32) (Bolger et al. 2014) with default settings (Table S1).

The full *A. kona* genome was initially assembled from 454 FLX + shotgun reads and Illumina paired-end reads as a *de novo* hybrid assembly using MIRA (v3.9.9) (Chevreux et al. 2004). Illumina mate-pair reads were then used for scaffolding using SSPACE (v2.0) (Boetzer et al. 2011). The assembly was filtered to remove redundant scaffolds, where redundancy was defined as scaffolds < 5 kb with a greater than 80% identity to another scaffold > 5 kb (Table S1a). The quality and completeness of the genome assembly was assessed using QUAST (Gurevich et al. 2013), and CEGMA (version 2.4) (Parra et al. 2007) and BUSCO (version 3.0) (Simão et al. 2015) (Tables S2a and S2b, respectively).

RNA Extraction and Transcriptome Sequencing

Cells were grown in Spiegel's liquid medium (Spiegel 1982) and total RNA was extracted using TRI Reagent LS (Sigma-Aldrich). Stranded mRNA libraries were constructed separately from total cultures as well as germinating spores, actively growing cells and aggregating cells (Fig. 1) using Illumina's Truseq RNA sample prep kit and sequenced on a HiSeq2000 illumina platform (2×100 bp, Pair-End). Adapter sequences and low quality bases were removed using Trimmomatic v.0.32 (Bolger et al. 2014) with default settings, and *de novo* transcriptome assemblies generated using Trinity (version 2014-07-17). The latter followed the protocol by Grabherr et al. 2011 with the --jaccard_clip option set to "yes" to minimize the fusion of transcripts in a relatively gene-compact genome.

For the three different life stages (Fig. 1) sequence quality was checked using FastQC (v0.11.8) (Andrews 2014), and the quality trimmed sequences were aligned to the genome assembly with STAR (version 2.5.2) (Dobin et al. 2013). Reads mapping to genes were counted using default settings in featureCounts program in R (Liao et al. 2014). As replicates were unavailable for each stage, GFOLD (v1.1.4) (Feng et al. 2012) was used to calculate fold change between growth and aggregation (gro v agg), and, aggregation and germination (agg v germ) (Table S10).

Gene prediction

Before gene prediction, low-complexity and interspersed repeat sequences were identified and masked using RepeatRunner (Smith et al. 2007) and RepeatMasker (www.repeatmasker.org/), the latter including novel and non-novel repeats from *Naegleria gruberi* (Fritz-Laylin et al. 2010). Genes were predicted using Augustus (version 2.7) (Stanke and Morgenstern 2005) and SNAP (November 2013 release) (Korf 2004) in the Maker pipeline (Cantarel et al. 2008) using parallel strands of evidence-based and *ab initio* gene prediction (Figure S10). Evidence-based gene prediction utilized both protein and transcriptome data to train the gene predictors. Protein evidence consisted of the UniProt/SwissProt database and the predicted proteome of *N. gruberi* (JGI release Naegr1). Transcriptome data consisted of pooled 454 and RNAseq data.

Ab initio gene prediction used the same input as above, together with the resulting "evidence-based" gene models and a set of 862 best gene models selected using the PASA package (Haas et al. 2003) based on four criteria: i) completeness (presence of start and stop codons), ii) physical distance from other genes (> 500 base pairs from any other predicted gene model), iii) non-redundant coding sequence (removing isoforms or close homologs), and iv) availability of RNAseq data for the full length of the gene.

A final round of gene prediction was then run using all raw input plus the results of both the evidence-based and *ab initio* builds to train the gene predictors and then passed through the MAKER pipeline again. This final step was performed in order to check the congruence of the predicted gene models from the evidence-based and *ab initio* strategies. The final result was four tracks of gene prediction: evidence-based gene predictions, RNAseq-based PASA predictions, *ab initio* predictions and combined *ab initio* and evidence-based gene model predictions. These were then curated manually to identify and resolve any conflicts (Figure S10).

Manual curation

The four gene prediction tracks were uploaded into Web Apollo (Lee et al. 2013) using the National Bioinformatics Infrastructure Sweden (NBIS) resources (https://www.nbis.se/). Each predicted gene was then checked by eye for congruence among the prediction methods. In cases of incongruence, all lines of evidence were re-checked, and the most reliable was selected. A gene model was considered reliable if it had start and stop codons, canonical intron splice sites (in most cases), > 500 nucleotide 5' and 3' flanking sequence (UTR), and full RNAseq coverage and/or at least one Pfam functional domain (Finn et al. 2014). Predicted genes with non-canonical splice sites were compared with RNAseq evidence and assembly coverage to check for possible errors and were otherwise kept. The predicted genes were checked for duplication using Python scripts. All gene duplicates with 100% identity at the nucleotide level were inspected manually by aligning their parent contigs using EMBOSS stretcher (Rice et al. 2000) and by checking their read coverage in Tablet (Milne et al. 2013). This was followed by a final check to identify and combine overlapping contigs and remove duplicate contig fragments.

Gene function annotation

Functional annotation of the fully curated gene models was carried out using the publicly available NBIS functional annotation pipeline (https://github.com/NBISweden/pipelines-nextflow). Each predicted gene model was searched at the deduced protein level using BLASTp against the UniProt/SwissProt reference database in order to retrieve gene and protein names and predicted protein functions. The predicted proteins were also passed through InterProScan (version 5.7–48) (Jones et al. 2014) in order to retrieve InterPro, Pfam (Finn et al. 2014) and Gene Ontology (GO) (Ashburner et al. 2000) data. These metadata were then parsed into the genome annotation files using Annotation Information Extractor (Annie) (Ooi et al. 2009) for viewing in WebApollo (Lee et al. 2013) (Figure S12). Gene ontology annotations for the predicted genes were also obtained using Blast2GO (Conesa et al. 2005) with default settings to confirm the consistency of the results. Protein translations of the curated gene models were also searched using BLASTp against the NCBI non-redundant protein database with an e-value cut-off of 1e-5. Metabolic pathways were annotated for both *Acrasis kona* and *Naegleria gruberi* using KEGG Orthology (KO) numbers using BlastKOALA (Kanehisa et al. 2016). This was also done for *Naegleria gruberi*. The resulting KO numbers were used as input to iPath (Letunic et al. 2008) to compare the two metabolic maps.

Gene families

Gene families were defined by all-vs-all BLASTp and clustering using OrthoMCL (Li et al. 2003). Clustering included proteomes from a broad taxonomic sampling of eukaryotes (Table S5). OrthoMCL was used with default settings with an e-value cut-off of 1e-10 for BLAST and an inflation value of 1.5 for MCL. The analysis was then repeated with ProteinOrtho (Lechner et al. 2011) with default settings (Table S6). The results were processed using Perl scripts to identify the numbers of genes in each family for each organism.

Horizontal gene transfer

Complete NCBI databases were downloaded from RefSeq (April 2016) for Metazoa, Fungi, Viridiplantae, Bacteria and Archaea and non-redundant protein databases for Amoebozoa, Stramenopila, Rhizaria, Alveolata, Rhodophyta and Excavata (Discoba + Metamonada). Transcriptomes for *Percolomonas cosmopolites* strains AE and WS (Heterolobosea, Discoba) were downloaded from iMicrobe (Keeling et al. 2014). Transcriptomes of *Andalucia godoyi* and *Seculomonas ecuadoriensis* (Jakobida, Discoba) were generated locally (He et al. 2016). Local individual BLAST databases were created for each of the major groups above, and the *Acrasis kona* predicted proteins were queried against each database with an e-value cut-off of 1e-10. The analyses were repeated separately for *Naegleria gruberi* (JGI release Naegr1) (Fritz-Laylin et al. 2010) and *Dictyostelium purpureum* (assembly v1.0) (Heidel et al. 2011) for comparison.

Acrasis kona proteins with hits only to Bacteria, Archaea or any single major eukaryotic group (“kingdoms”, Adl et al. 2019) other than Excavata were screened for potential signal peptides using the SignalP (version 4.0) (Petersen et al. 2011) and Phobius webserver (Käll et al. 2007). Functional categorization of these proteins was done using GoFeat (Araujo et al. 2018). All non-novel *A. kona*

proteins with non-excavate top hits and total hits (e-value < 1e-35) to three or more major eukaryote groups were analyzed phylogenetically using a bioinformatic pipeline as follows. i) Each protein was searched against the local NCBI database for each major group separately, and all the hits with e-value < 1e-35 were retrieved. ii) Sequences were aligned using MUSCLE (Edgar 2004) in AliView (Larsson 2014), and alignments were trimmed using the heuristic (automated1) algorithm in trimAL (Capella-Gutiérrez et al. 2009). iii) Maximum likelihood trees were constructed using IQTree (Chernomor et al. 2016) with standard model selection (-m TEST) and the SH-aLRT and ultrafast bootstrap (1000 replicates) tests. iv) Evolutionary relationships in each resulting tree were identified using SICLE (DeBlasio and Wisecaver 2016).

Developmental transcriptome analysis

Accessions with differential expression (DE) during *Acrasis kona* development were annotated based on a consensus of top BLASTp hits to 1) all GenBank taxa (nr database), 2) human (SwissProt) associated annotation and 3) *D. discoideum AX4* (RefSeq) and associated Dictybase annotation (http://dictybase.org/, Fey et al. 2019). Accessions were manually clustered into categories based on their function predictions, and also automatically annotated and clustered using BlastKOALA or, for larger data sets, GhostKOALA, with an unspecified target taxon (https://www.expasy.org/). Koala annotations per functional categories were then tabulated after automated removal of duplicates. Transmembrane protein structure and orientation were predicted using TMHMM 2.0 (Krogh et al. 2001) and potential G-protein coupled receptors by GPCRHMM (Wistrand et al. 2006).

The proteasome, exosome and kinetochore were predicted for diverse eukaryotes by BLASTp at NCBI using genus- or species-level taxid limits and human RefSeq queries. Homologous sequences for *Acrasis kona* were first identified by local standalone BLASTp and then confirmed by BLASTp against the human RefSeq database. For the autophagosome, a universal protein set was first predicted by assembling a consensus of all GenBank autophagy annotated accessions for human, *Saccharomyces cerevisiae S288C, Schizosaccharomyces pombe* and *Dictyostyelium discoideum AX4*. Sequences for diverse eukaryotes were then identified by BLASTp using taxid limits and *D. discoideum* AX4 queries. Mean read counts for the normalized reads for each replicate for *D. discoideum* AX4 developmental genes were retrieved from Santhanam et al (2015). Differential expression was calculated for zero, two and five hours after food was depleted (starvation) using $\log^2$-fold change between time points.

Phylogenetic analyses of potential aggregation-specific horizontal gene transfers were conducted at the amino acid level using sequences identified by BLASTp against GenBank nr. Sequences were aligned using MUSCLE (v3.8.32) (Edgar 2004) as implemented in AliView (Larsson 2014) and trimmed using trimAl (automated1 parameter) (Capella-Gutiérrez et al. 2009). Trees were constructed using RAxML with the LG + gamma substitution model and bootstrapping with automated stopping (Stamatakis 2006) on the CIPRES webserver (Miller et al. 2010).

Declarations

Acknowledgements:

We thank Robert Insall for the *Acrasis* movie and Wei Miao for generously funding the initial genome sequencing. Thanks also to Marc P. Höppner, Jacques Dainat and the Uppsala Science for Life (SciLife) lab for technical support, and the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX), the Swedish Bioinformatics Infrastructure for Life Sciences (BILs) and the Cyberinfrastructure for Phylogenetic Research (CIPRES) for use of computational resources. Organelle schematics for the graphic abstract are courtesy of BioRender. This work was supported by grants from the Swedish Research Council (2017-04351 Vetenskapsrådet) and the National Natural Science Foundation of China. MWB is supported by the United States National Science Foundation (NSF) Division of Environmental Biology (DEB) grant 2100888 (http://www.nsf.gov).

Funding:

Swedish Research Council (Vetenskapsrådet) VR 2017-04351 (SLB).

Uppsala University, Department of Organismal Biology (CAJ)

Author contributions:

Conceptualization: SLB

Methodology: SS, CJF, SLB

Formal Analysis: SS, CJF, SLB

Data Curation: SS, CJF, SLB

Resources: MWB, SLB

Data interpretation: SS, CJF, MWB, SLB

Visualization: SS, CJF, SLB

Funding acquisition: SLB

Project administration: SLB

Writing: SS, CJF, SLB

Competing interests:

Authors declare that they have no competing interests.

Data and materials availability:

All sequence data associated with this study are publicly available under the GenBank Bioproject PRJNA850442. This includes the genome assembly (accession number GCA_026419775.1) and all transcriptome data (SRA file accession numbers SRR22861965, SRX19285604, SRX19285605, and SRX19285603). *Acrasis kona* strain MYA-3509 (formerly *Acrasis rosea*) is available from the American Type Culture Collection (ATCC:MYA-3509).

References

1. Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F, Cárdenas P, Čepička I, Chistyakova L, Del Campo J, Dunthorn M, Edvardsen B, Eglit Y, Guillou L, Hampl V, Heiss AA, Hoppenrath M, James TY, Karnkowska A, Karpov S, Kim E, Kolisko M, Kudryavtsev A, Lahr DJG, Lara E, Le Gall L, Lynn DH, Mann DG, Massana R, Mitchell EAD, Morrow C, Park JS, Pawlowski JW, Powell MJ, Richter DJ, Rueckert S, Shadwick L, Shimano S, Spiegel FW, Torruella G, Youssef N, Zlatogursky V, Zhang Q. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. J Eukaryot Microbiol. 66(1):4–119 (2019).
2. Al Jewari C, Baldauf SL. Conflict over the eukaryote root resides in strong outliers, mosaics and missing data sensitivity of site-specific (CAT) mixture models. Syst Biol syac029 (2022).
3. Andrews S. A quality control tool for high throughput sequence data. FastQC http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2014).
4. Araujo FA, Barh D, Silva A, Guimarães L, Ramos RTJ. GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. Sci Rep. 8(1):1794 (2018).
5. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 25(1):25–9 (2000).
6. Baldauf SL, Romeralo M, Fiz-Palacios O, Heidari N. A Deep Hidden Diversity of Dictyostelia. Protist 169(1):64–78 (2018).
7. Baldauf SL, Strassmann JE. Dictyostelia. In: Archibald J., Simpson A., Slamovits C. (eds) Handbook of the Protists. Springer, Cham (2017).
8. Bard JAM, Goodall EA, Greene ER, Jonsson E, Dong KC, Martin A. Structure and Function of the 26S Proteasome. Annu Rev Biochem. 87:697–724 (2018).
9. Bakthavatsalam D, Gomer RH. The secreted proteome profile of developing Dictyostelium discoideum cells. Proteomics 10(13):2556–9 (2010).
10. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27(4):578–9 (2011).
11. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–20 (2014).

12. Brown MW, Kolisko M, Silberman JD, Roger AJ. Aggregative multicellularity evolved independently in the eukaryotic supergroup Rhizaria. Curr Biol. 22(12):1123–7 (2012b).
13. Brown MW, Silberman JD, Spiegel FW. A morphologically simple species of Acrasis (Heterolobosea, Excavata), Acrasis helenhemmesae n. sp. J Eukaryot Microbiol. 57(4):346–53 (2010).
14. Brown MW, Silberman JD, Spiegel FW. "Slime molds" among the Tubulinea (Amoebozoa): molecular systematics and taxonomy of Copromyxa. Protist. 162(2):277–87 (2011).
15. Brown MW, Silberman JD, Spiegel FW. A contemporary evaluation of the acrasids (Acrasidae, Heterolobosea, Excavata). Eur J Protistol 48(2):103–23 (2012a).
16. Brown MW, Spiegel FW, Silberman JD, 2009. Phylogeny of the "Forgotten" Cellular Slime Mold, Fonticula alba, Reveals a Key Evolutionary Branch within Opisthokonta. Mol Biol Evol 26: 2699–2709 (2009).
17. Budenholzer L, Cheng CL, Li Y, Hochstrasser M. Proteasome Structure and Assembly. J Mol Biol. 429(22):3500–3524 (2017).
18. Burki F, Roger AJ, Brown MW, Simpson AGB. The New Tree of Eukaryotes. Trends Ecol Evol. 35(1):43–55 (2020).
19. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18(1):188–96 (2008).
20. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972–3 (2009).
21. Chen G, Kuspa A. Prespore cell fate bias in G1 phase of the cell cycle in Dictyostelium discoideum. Eukaryot Cell 4(10):1755–64 (2005).
22. Cheng X, Ma X, Ding X, Li L, Jiang X, Shen Z, Chen S, Liu W, Gong W, Sun Q. Pacer Mediates the Function of Class III PI3K and HOPS Complexes in Autophagosome Maturation by Engaging Stx17. Mol Cell. 65(6):1029–1043.e5 (2017)
23. Chernomor O, von Haeseler A, Minh BQ. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. Syst Biol. 65(6):997–1008 (2016).
24. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WE, Wetter T, Suhai S. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 14(6):1147–59 (2004).
25. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–6 (2005).
26. DeBlasio DF, Wisecaver JH. SICLE: a high-throughput tool for extracting evolutionary relationships from phylogenetic trees. PeerJ. 4:e2359 (2016).
27. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1):15–21 (2013).

28. Doolittle WF. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet. 14(8):307–11 (1998).
29. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–7 (2004).
30. Eichinger L, Pachebat JA, Glöckner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P, Fey P, Pilcher K, Chen G, Saunders D, Sodergren E, Davis P, Kerhornou A, Nie X, Hall N, Anjard C, Hemphill L, Bason N, Farbrother P, Desany B, Just E, Morio T, Rost R, Churcher C, Cooper J, Haydock S, van Driessche N, Cronin A, Goodhead I, Muzny D, Mourier T, Pain A, Lu M, Harper D, Lindsay R, Hauser H, James K, Quiles M, Madan Babu M, Saito T, Buchrieser C, Wardroper A, Felder M, Thangavelu M, Johnson D, Knights A, Loulseged H, Mungall K, Oliver K, Price C, Quail MA, Urushihara H, Hernandez J, Rabbinowitsch E, Steffen D, Sanders M, Ma J, Kohara Y, Sharp S, Simmonds M, Spiegler S, Tivey A, Sugano S, White B, Walker D, Woodward J, Winckler T, Tanaka Y, Shaulsky G, Schleicher M, Weinstock G, Rosenthal A, Cox EC, Chisholm RL, Gibbs R, Loomis WF, Platzer M, Kay RR, Williams J, Dear PH, Noegel AA, Barrell B, Kuspa A. The genome of the social amoeba Dictyostelium discoideum. Nature 435(7038):43–57 (2005).
31. Feng J, Meyer CA, Wang Q, Liu JS, Shirley Liu X, Zhang Y. GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. Bioinformatics 28(21):2782–8 (2012).
32. Fernández R, Gabaldón T. Gene gain and loss across the metazoan tree of life. Nat Ecol Evol. 4(4):524–533 (2020).
33. Fey P, Dodson RJ, Basu S, Hartline EC, Chisholm RL. dictyBase and the Dicty Stock Center (version 2.0) - a progress report. Int J Dev Biol. 63(8-9-10):563–572 (2019).
34. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. Pfam: the protein families database. Nucleic Acids Res. 42(Database issue):D222-30 (2014).
35. Fischer S, Eichinger L. Dictyostelium discoideum and autophagy - a perfect pair. Int J Dev Biol. 63(8-9-10):485–495 (2019).
36. Fiz-Palacios O, Romeralo M, Ahmadzadeh A, Weststrand S, Ahlberg PE, Baldauf S. Did terrestrial diversification of amoebas (amoebozoa) occur in synchrony with land plants? PLoS One 8(9):e74374 (2013).
37. Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, Shu S, Neupane R, Cipriano M, Mancuso J, Tu H, Salamov A, Lindquist E, Shapiro H, Lucas S, Grigoriev IV, Cande WZ, Fulton C, Rokhsar DS, Dawson SC. The genome of Naegleria gruberi illuminates early eukaryotic versatility. Cell 140(5):631–42 (2010).
38. Fu CJ, Sheikh S, Miao W, Andersson SG, Baldauf SL. Missing genes, multiple ORFs, and C-to-U type RNA editing in Acrasis kona (Heterolobosea, Excavata) mitochondrial DNA. Genome Biol Evol. 6(9):2240–57 (2014).

39. Gladyshev EA, Arkhipova IR. A widespread class of reverse transcriptase-related cellular genes. Proc Natl Acad Sci USA 108(51):20311–6 (2011).
40. Glöckner G, Lawal HM, Felder M, Singh R, Singer G, Weijer CJ, Schaap P. The multicellularity genes of dictyostelid social amoebas. Nat Commun. 7:12085 (2016).
41. Gomer RH, Jang W, Brazill D. Cell density sensing and size determination. Dev Growth Differ 53(4):482–94 (2011).
42. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29(7):644–52 (2011).
43. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–5 (2013).
44. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, Salzberg SL, White O. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31(19):5654–66 (2003).
45. He D, Fu CJ, Baldauf SL. Multiple origins of eukaryotic cox15 suggest horizontal gene transfer from bacteria to jakobid mitochondrial DNA. Mol Biol Evol. 33(1):122–33 (2016).
46. Heidel AJ, Lawal HM, Felder M, Schilde C, Helps NR, Tunggal B, Rivero F, John U, Schleicher M, Eichinger L, Platzer M, Noegel AA, Schaap P, Glöckner G. Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. Genome Res. 21(11):1882–91 (2011).
47. Helassa N, Antonyuk SV, Lian LY, Haynes LP, Burgoyne RD. Biophysical and functional characterization of hippocalcin mutants responsible for human dystonia. Hum Mol Genet. 26(13):2426–2435 (2017).
48. Hellstén M, Roos UP. The actomyosin cytoskeleton of amoebae of the cellular slime molds acrasis rosea and protostelium mycophaga: structure, biochemical properties, and function. Fungal Genet Biol. 24(1–2):123–45 (1998).
49. Huber RJ, O'Day DH. Proteomic profiling of the extracellular matrix (slime sheath) of Dictyostelium discoideum. Proteomics 15(19):3315–9 (2015).
50. Jaiswal P, Kimmel AR. mTORC1/AMPK responses define a core gene set for developmental cell fate switching. BMC Biology 17:58 (2019).
51. Januszyk K, Lima CD. The eukaryotic RNA exosome. Curr Opin Struct Biol. 24:132–40 (2014).
52. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–40 (2014).
53. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 428(4):726–731 (2016).

54. Kawabe Y, Du Q, Schilde C, Schaap P. Evolution of multicellularity in Dictyostelia. Int J Dev Biol. 63(8-9-10):359–369 (2019).
55. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B, Bidle KD, Cameron CT, Campbell L, Caron DA, Cattolico RA, Collier JL, Coyne K, Davy SK, Deschamps P, Dyhrman ST, Edvardsen B, Gates RD, Gobler CJ, Greenwood SJ, Guida SM, Jacobi JL, Jakobsen KS, James ER, Jenkins B, John U, Johnson MD, Juhl AR, Kamp A, Katz LA, Kiene R, Kudryavtsev A, Leander BS, Lin S, Lovejoy C, Lynn D, Marchetti A, McManus G, Nedelcu AM, Menden-Deuer S, Miceli C, Mock T, Montresor M, Moran MA, Murray S, Nadathur G, Nagai S, Ngam PB, Palenik B, Pawlowski J, Petroni G, Piganeau G, Posewitz MC, Rengefors K, Romano G, Rumpho ME, Rynearson T, Schilling KB, Schroeder DC, Simpson AG, Slamovits CH, Smith DR, Smith GJ, Smith SR, Sosik HM, Stief P, Theriot E, Twary SN, Umale PE, Vaulot D, Wawrik B, Wheeler GL, Wilson WH, Xu Y, Zingone A, Worden AZ. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 12(6):e1001889 (2014).
56. Kleerebezem M, Quadri LE, Kuipers OP, de Vos WM. Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria. Mol Microbiol. 24(5):895–904 (1997).
57. Korf I. Gene finding in novel genomes. BMC Bioinformatics 5:59 (2004).
58. Kowalinski E, Kögel A, Ebert J, Reichelt P, Stegmann E, Habermann B, Conti E. Structure of a Cytoplasmic 11-Subunit RNA Exosome Complex. Mol Cell 63(1):125–34 (2016).
59. Kranenburg O, Moolenaar WH. Ras-MAP kinase signaling by lysophosphatidic acid and other G protein-coupled receptor agonists. Oncogene. 20(13):1540–6 (2001).
60. Kriebel PW, Parent CA. Adenylyl cyclase expression and regulation during the differentiation of Dictyostelium discoideum. IUBMB Life. 56(9):541–6 (2004).
61. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 305(3):567–80 (2001).
62. Kuo CJ, Hansen M, Troemel E. Autophagy and innate immunity: Insights from invertebrate model organisms. Autophagy 14(2):233–242 (2018).
63. Käll L, Krogh A, Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. Nucleic Acids Res. 35(Web Server issue):W429-32 (2007).
64. Łabno A, Tomecki R, Dziembowski A. Cytoplasmic RNA decay pathways - Enzymes and mechanisms. Biochim Biophys Acta. 1863(12):3125–3147 (2016).
65. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30(22):3276–8 (2014).
66. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics 12:124 (2011).

67. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE. Web Apollo: a web-based genomic annotation editing platform. Genome Biol. 14(8):R93 (2013).
68. Letunic I, Yamada T, Kanehisa M, Bork P. iPath: interactive exploration of biochemical pathways and networks. Trends Biochem Sci. 33(3):101–3 (2008).
69. Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13(9):2178–89 (2003).
70. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30(7):923–30 (2014).
71. Loomis WF. Cell signaling during development of Dictyostelium. Dev Biol. 391(1):1–16 (2014).
72. Marie B, Zanella-Cléon I, Guichard N, Becchi M, Marin F. Novel proteins from the calcifying shell matrix of the Pacific oyster Crassostrea gigas. Mar Biotechnol (NY). 13(6):1159–68 (2011).
73. Mazur M, Wojciechowska D, Sitkiewicz E, Malinowska A, Świderska B, Kmita H, Wojtkowska M. Mitochondrial Processes during Early Development of *Dictyostelium discoideum*: From Bioenergetic to Proteomic Studies. Genes (Basel). 12(5):638 (2021).
74. Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010) "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1–8.
75. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D. Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform. 14(2):193–202 (2013).
76. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 39:121–52 (2005)
77. Olive LS. The Mycetozoans. Academic Press, Inc (1975).
78. Ooi HS, Kwo CY, Wildpaner M, Sirota FL, Eisenhaber B, Maurer-Stroh S, Wong WC, Schleiffer A, Eisenhaber F, Schneider G. ANNIE: integrated de novo protein sequence annotation. Nucleic Acids Res. 37(Web Server issue):W435-40 (2009).
79. Page FC, Blanton RL. The Heterolobosea (Sarcodina: Rhizopoda) a new class uniting the Schizopyrenida and the Acrasidae (Acrasida). Protistologica 21(1): 121–132 (1985).
80. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23(9):1061–7 (2007).
81. Pánek T, Simpson AGB, Brown MW, Dyer BD. Heterolobosea. In: Archibald J, Simpson A, Slamovits C. (eds) Handbook of the Protists. Springer, Cham (2017).
82. Patel AL, Shvartsman SY. Outstanding questions in developmental ERK signaling. Development. 145(14):dev143818 (2018).
83. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10):785–6 (2011).

84. Pires-daSilva A, Sommer RJ. The evolution of signaling pathways in animal development. Nat Rev Genet. 4(1):39–49 (2003).
85. Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishii I, Ferris P, Kuo A, Mitros T, Fritz-Laylin LK, Hellsten U, Chapman J, Simakov O, Rensing SA, Terry A, Pangilinan J, Kapitonov V, Jurka J, Salamov A, Shapiro H, Schmutz J, Grimwood J, Lindquist E, Lucas S, Grigoriev IV, Schmitt R, Kirk D, Rokhsar DS. Genomic analysis of organismal complexity in the multicellular green alga Volvox carteri. Science. 329:223–6 (2010).
86. Raper K. 1985. *The Dictyostelids.* Princeton University Press.
87. Reinhardt JD. The Effects of Light on the Development of the Cellular Slime Mold Acrasis rosea. American Journal of Botany 55(1):77–86 (1968).
88. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16(6):276–7 (2000).
89. Roger AJ, Smith MW, Doolittle RF, Doolittle WF. Evidence for the Heterolobosea from phylogenetic analysis of genes encoding glyceraldehyde-3-phosphate dehydrogenase. J Eukaryot Microbiol. 43(6):475–85 (1996).
90. Romeralo M, Skiba A, Gonzalez-Voyer A, Schilde C, Lawal H, Kedziora S, Cavender JC, Glöckner G, Urushihara H, Schaap P. Analysis of phenotypic evolution in Dictyostelia highlights developmental plasticity as a likely consequence of colonial multicellularity. Proc Biol Sci. 280(1764):20130976 (2013).
91. Santhanam B, Cai H, Devreotes PN, Shaulsky G, Katoh-Kurasawa M. The GATA transcription factor GtaC regulates early developmental gene expression dynamics in Dictyostelium. Nat Commun. 6:7551 (2015).
92. Schaap P, Schilde C. Encystation: the most prevalent and under-investigated differentiation pathway of eukaryotes. Microbiology (Reading). 164(5):727–739. (2018).
93. Scherer A, Kuhl S, Wessels D, Lusche DF, Raisley B, Soll DR. Ca2 + chemotaxis in Dictyostelium discoideum. J Cell Sci. 123(Pt 21):3756–67 (2010).
94. Sebé-Pedrós A, Irimia M, Del Campo J, Parra-Acero H, Russ C, Nusbaum C, Blencowe BJ, Ruiz-Trillo I. Regulated aggregative multicellularity in a close unicellular relative of metazoa. Elife 2:e01287 (2013).
95. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–2 (2015).
96. Siu CH, Harris TJ, Wang J, Wong E. Regulation of cell-cell adhesion during Dictyostelium development. Semin Cell Dev Biol. 15:633–41 (2004).
97. Smith CD, Edgar RC, Yandell MD, Smith DR, Celniker SE, Myers EW, Karpen GH. Improved repeat identification and masking in Dipterans. Gene. 389(1):1–9 (2007).
98. Spiegel FW. The ultrastructure of the trophic cells of the protostelid Planoprotostelium aurantium. Protoplasma 113:165–177 (1982).

99. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatic 22(21):2688–90 (2006).
100. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33(Web Server issue):W465-7 (2005).
101. Strassert JFH, Irisarri I, Williams TA, Burki F, A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. Nat Commun. 12:1879 (2021).
102. Strassmann JE, Gilbert OM, Queller DC. Kin discrimination and cooperation in microbes. Annu Rev Microbiol. 65:349–67 (2011).
103. Sugimoto H, Endoh H. Analysis of fruiting body development in the aggregative ciliate Sorogena stoianovitchae (Ciliophora, Colpodea). J Eukaryot Microbiol. 53:96–102 (2006).
104. Tice AK, Silberman JD, Walthall AC, Le KN, Spiegel FW, Brown MW. Sorodiplophrys stercorea: Another Novel Lineage of Sorocarpic Multicellularity. J Eukaryot Microbiol. 63(5):623–8 (2016).
105. Tromer EC, van Hooff JJE, Kops GJPL, Snel B. Mosaic origin of the eukaryotic kinetochore. Proc Natl Acad Sci USA 116(26):12873–12882 (2019).
106. van Hooff JJ, Tromer E, van Wijk LM, Snel B, Kops GJ. Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. EMBO Rep. 18(9):1559–1571 (2017).
107. van Tieghem MP. Sur Quelques Myxomycetes A Plasmode Agrege. Bulletin de la Société botanique de France 27: 317–322 (1880).
108. Wang N, Söderbom F, Anjard C, Shaulsky G, Loomis WF. SDF-2 induction of terminal differentiation in Dictyostelium discoideum is mediated by the membrane-spanning sensor kinase DhkA. Mol Cell Biol. 19(7):4750–6 (1999).
109. Wistrand M, Käll L, Sonnhammer EL. A general model of G protein-coupled receptor sequences and its application to detect remote homologs. Protein Sci. 15(3):509–21 (2006).

Table

Table 1. Genome statistics for *A. krona* and selected species

Species	Genome size (mB)	%GC	Protein-coding loci	% coding	% genes with introns	Introns per gene	Median intron length
Acrasis kona	44	38	15868	45.5	48	1.1	40
Naegleria gruberi	41	57	15727	57.8	36	0.7	60
Homo sapiens	2851	41	23328	1.2	83	7.8	20383
Neurospora crassa	40	54	10107	36.4	80	1.7	72
Dictyostelium discoideum	34	22	13574	62.2	68	1.3	236
Arabidopsis thaliana	140.1	36	26541	23.7	80	4.4	55
Thalassiosira pseudonana	34.5	47	11242	ND	ND	1.4	ND
Saccharomyces cerevisiae	12	38	6039	73	ND	1	141

Figures

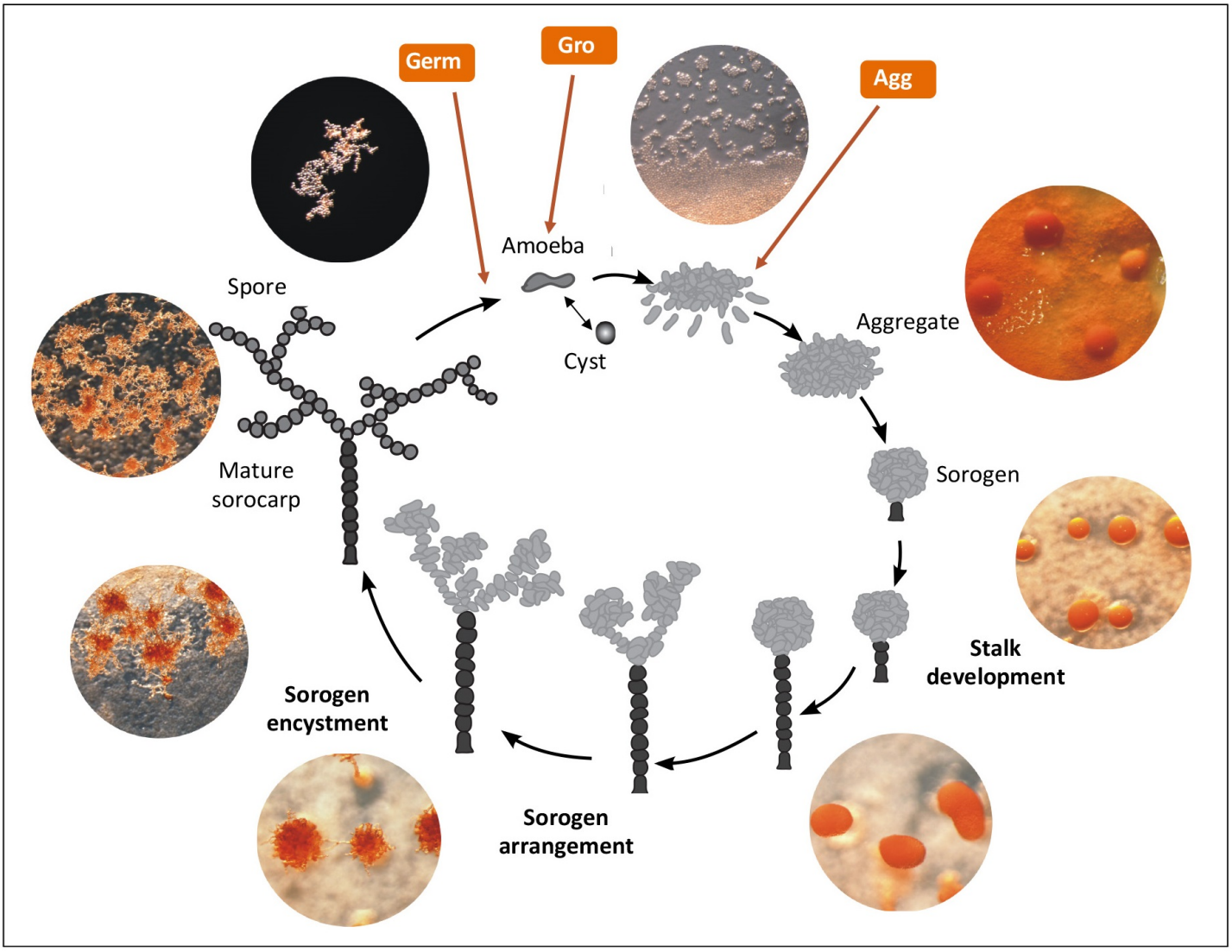


Figure 1

The Acrasis kona life cycle. The acrasid life cycle alternates between a single-celled (asocial) feeding stage and a multicellular (social) dispersal stage. Development begins with aggregation, where cells migrate together to form a ball of cells (sorogen) surrounded by an extracellular matrix (slime sheath). The basal cells then encyst to form a stalk, gradually lifting the remaining cell mass above the substrate. With the stalk complete, the aerial cells align into chains and then encyst en mass, creating a mature multicellular fruiting body (sorocarp). Red arrows indicate the three life cycle stage transcriptomes: Germ (spore germination), Gro (growth), and Agg (aggregation).

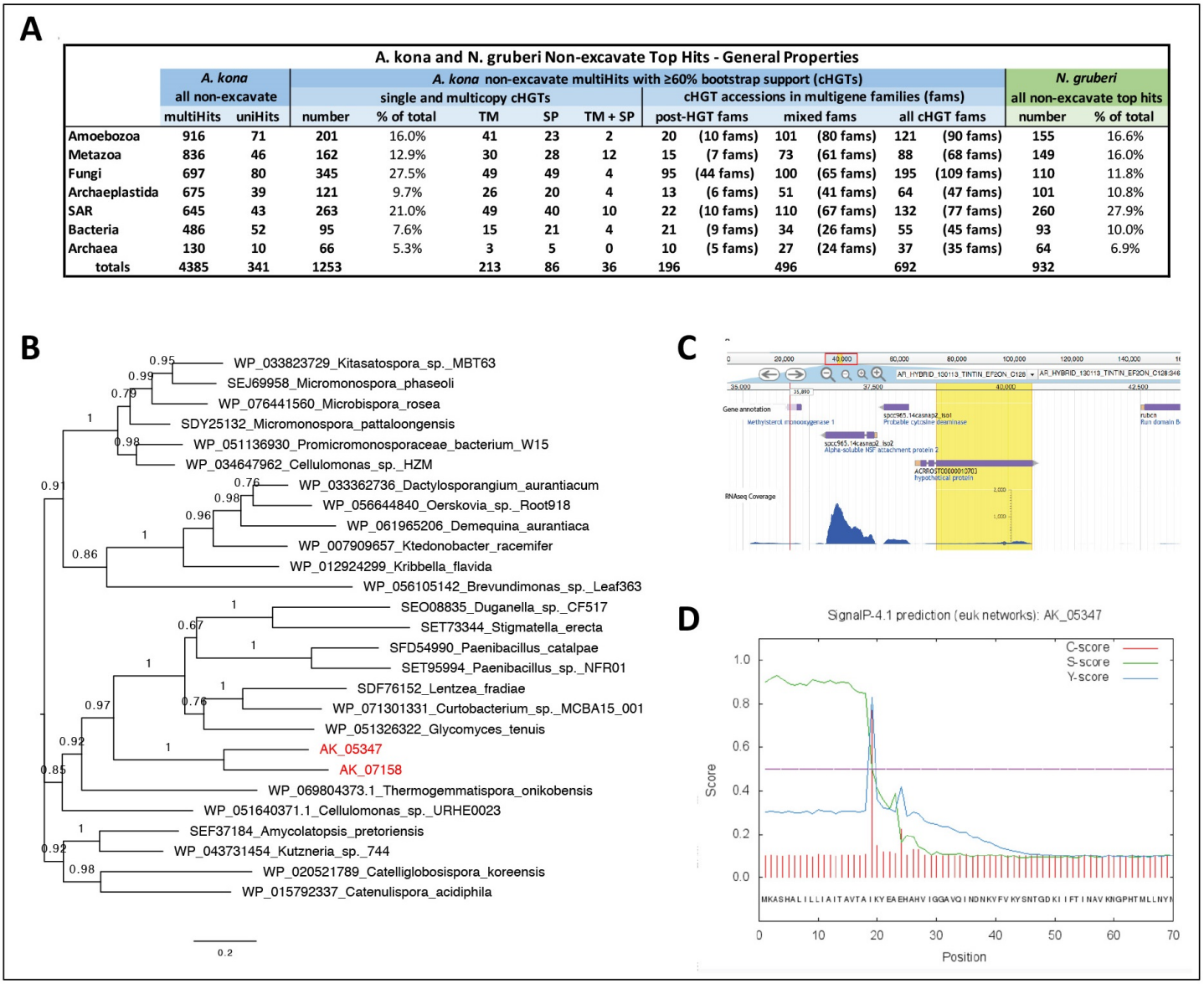


Figure 2

Horizontal gene transfer (HGT) in *Acrasis kona* including A) summary statistics and B-D) an example of HGT from Bacteria with addition of a signal sequence and introns. A) Total numbers of predicted proteins with non-Excavata top hits (BLASTp) are shown, organized by top hit taxonomy, for *Acrasis kona* (far left) and *Naegleria gruberi* (far right). *A. kona* hits are further classified as present in one (uniHits) or multiple (multiHits) non-excavate taxon groups. *A. kona* multiHits with >60% bootstrap support (confirmed HGTs or cHGTs) were further screened for signal peptides (SP), transmembrane domains (TM), and gene families (fams). Monophyletic multicopy cHGTs are classified as “post-HGT fams” - a single transfer followed by duplication, or “mixed fams” - multiple transfers or a mixture of horizontal and vertical transmission (full details in Table S8). B). A bacterial HGT followed by duplication (post-HGT fam), shown by maximum likelihood phylogeny of a 503 amino acid alignment. C) Both *A. kona* sequences carry two introns (shown for AK_05347), D) including an intron at the junction between the signal peptide and mature protein coding sequences.

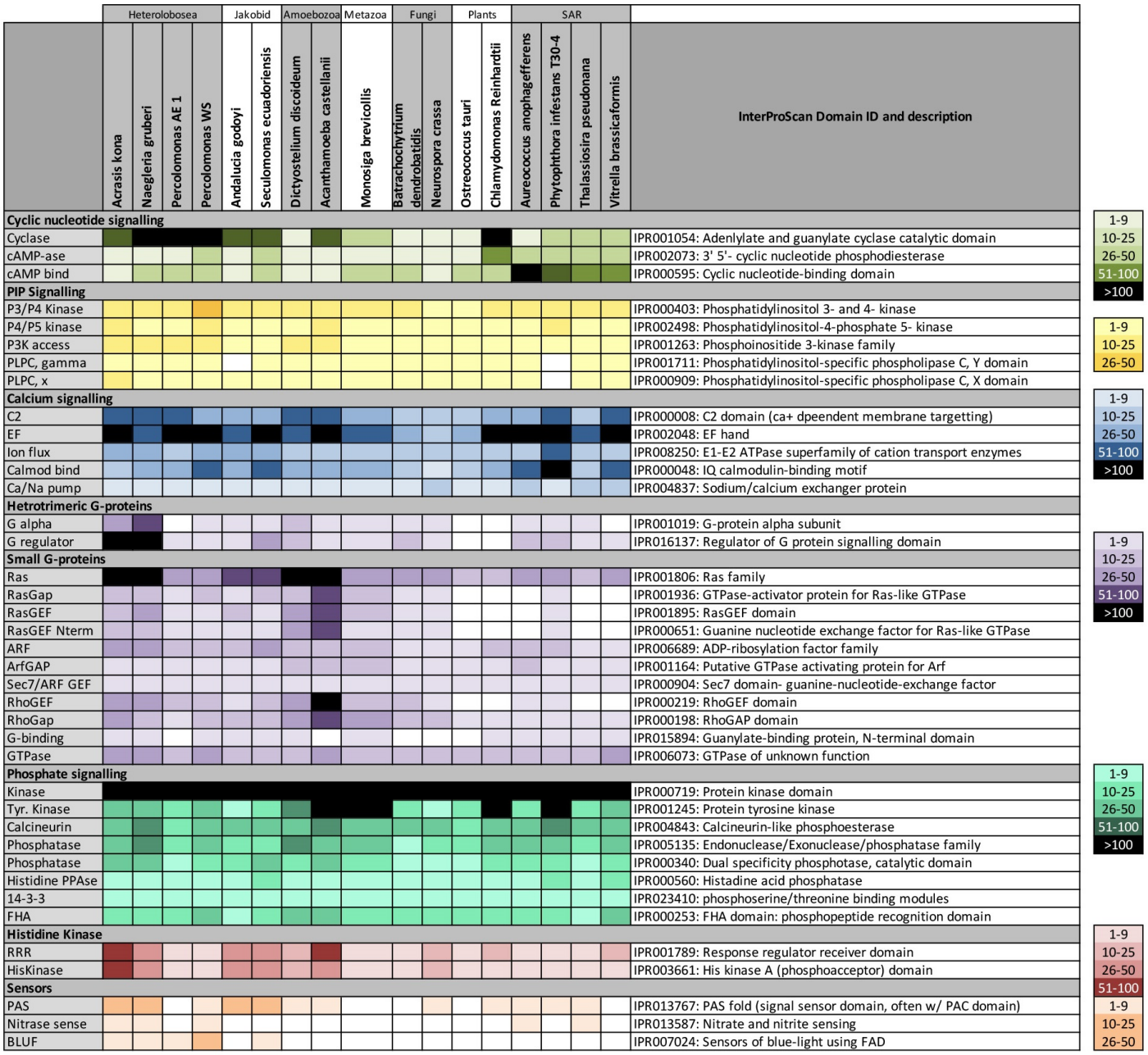


Figure 3

Signaling domains in *Acrasis kona* and other eukaryotes. Signaling domains on the left were identified by InterProScan (Jones et al. 2014) for *A. kona* and diverse other eukaryotes including five additional representatives of kingdom Discoba: two Jakobida (*Seculamonas equadoriensis* and *Andalucia godoyi*) and three distantly related Heterolobosea (*Naegleria gruberi*, *Percolomonas* "strain AE1" and *Percolomonas* "strain WS"). InterProScan IDs and descriptions of each domain are shown in the righthand column. The numbers of domains detected for each taxon are given in Table S9 and indicated in the figure by color intensity according to the keys at the far right.

A

category	*A. kona* 0-5 hr development				*Ddi* AX4 0-5 hr development		*A. kona* germination				*A. kona* full proteome	
	Aggup		Aggdn		Aggup	Aggdn	Germup		Germdn			
all	449		447		2762	1575	1627		1993		15868	
novel	94	(20.9%)	212	(47.4%)			488	(29.9%)	423	(21.2%)	4987	(31.4%)
Multicopy:	acc's (OGs)						acc's (OGs)				acc's (OGs)	
Ako	190	(161 OGs)	195	(175 OGs)			877	(558 OGs)	948	(656 OGs)	7766	(2728 OGs)
Ngr		(35 OGs)		(19 OGs)				(102 OGs)		(132 OGs)		(359 OGs)
all euks		(88 OGs)		(32 OGs)				(161 OGs)		(232 OGs)		(794 OGs)
cHGT	35		34				118		202		1594	
TM +/- SignalP	98	(21.8%)	126	(28.2%)			520	(32.0%)	647	(32.5%)	4479	(28.2%)

B

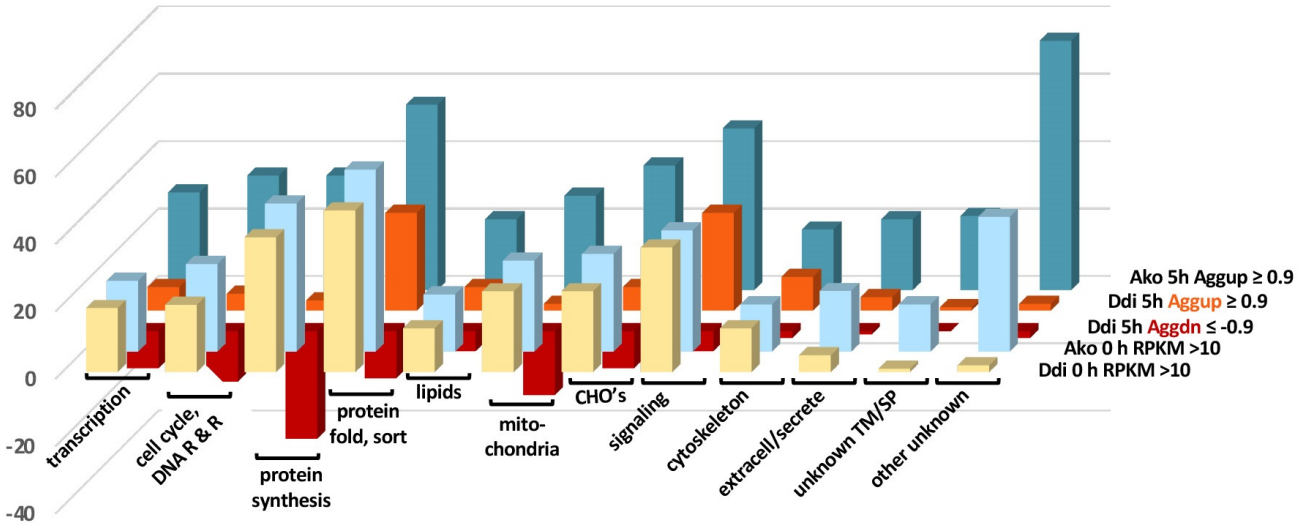


Figure 4

***Acrasis kona* developmental genes (A) and their functional profile compared to homologs in *Dictyostelium discoideum* AX4 (B).** A) Numbers of *A. kona* accessions with substantial differential expression (SDE) between growth and aggregation (Aggup, Aggdn) and aggregation and germination (Germup, Germdn) are summarized from Table S10. Numbers of SDE accessions are compared to the full *A. kona* proteome for novelty and membership in multiprotein families (Multicopy), including numbers of accessions (accs) and orthology groups (OGs) involved, as well as confirmed horizontal transfer (cHGT) and predicted signal peptides and/or transmembrane domains (Tm, SignalP). OGs are further divided into ones shared with *Naegleria gruberi* (Ngr) or a broad sampling of eukaryotes (all euks). (B) Functional profiles are shown for the 449 *A. kona* Aggup accessions and their Ddi AX4 homologs. Bar height corresponds to number of accessions with RPKM >10 during active growth (0h) or differentially expressed between growth and development (0_5h, DE_{Log2} ≥0.9 or ≤-0.9). Full *A. kona* annotations and functional clustering are given in Table S11. Category abbreviations are as follows: DNA replication and repair (DNA R&R), protein modification, folding and sorting (protein fold, sort), lipid chemistry (lipids), carbohydrate chemistry (CHO's), extracellular or secreted (extracell/secrete), predicted transmembrane domains and/or signal peptides (TM/SP). Differentially expression of Ddi AX4 genes was calculated using data from Santhanam et al. 2015 and employing the same criteria as for *A. kona* (Table S11).

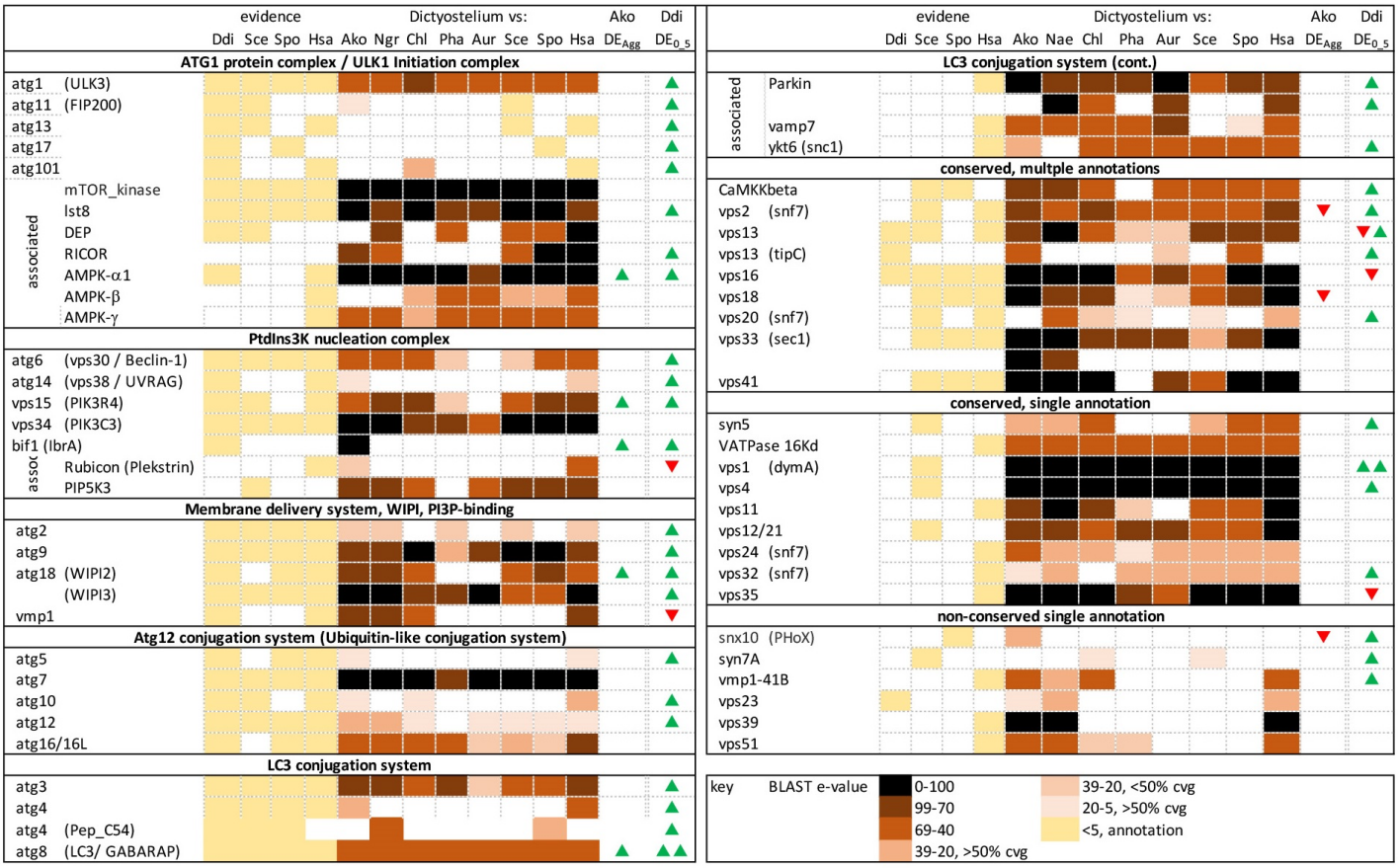


Figure 5

A universal predicted autophagosome and its expression in *Acrasis kona* and *Dictyostelium discoideum*. A universal autophagosome was predicted based on annotation in *Homo sapiens*, *Saccharomyces cerevisiae S288C*, *Schizosaccharomyces pombe* and *Dictyostelium discoideum AX4* (Ddi AX4). Homologs were then identified in other eukaryotes by BLASTp using Ddi AX4 protein sequences as queries. Levels of sequence similarity are indicated by color intensity (key at bottom right); BLASTp hits with e-values above e^{-39} were scored as present only if their BLASTp alignments included >50% query coverage (cvg). Increased (DE >0.9) or decreased (DE <-0.9) gene expression is indicated by green and red arrows, respectively, for aggregating *A. kona* (Ako DE_{Agg}) and Ddi AX4 (Ddi DE_{0_5}) (Table S11). Taxon names are denoted as Ako (*A. kona*), Aur (*Aureococcus* spp.), Chl (*Chlorella* spp.), Ddi (Ddi AX4), Has (*H. sapiens*), Ngr (*Naegleria* spp.), Pha (*Phaeodactylum* spp.), Sce (*S. cerevisiae*), Spo (*S. pombe*). Full details in Table S12

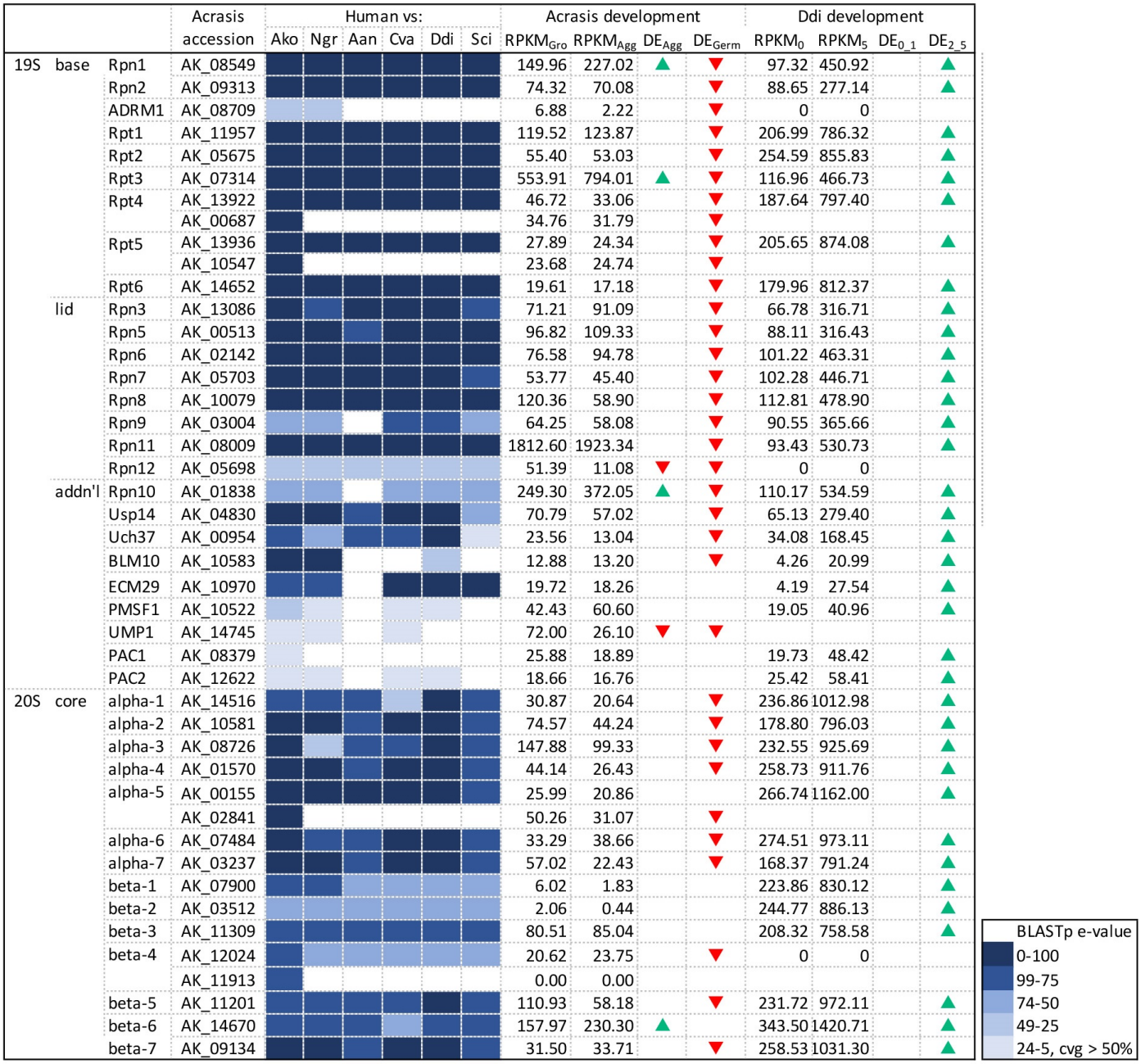
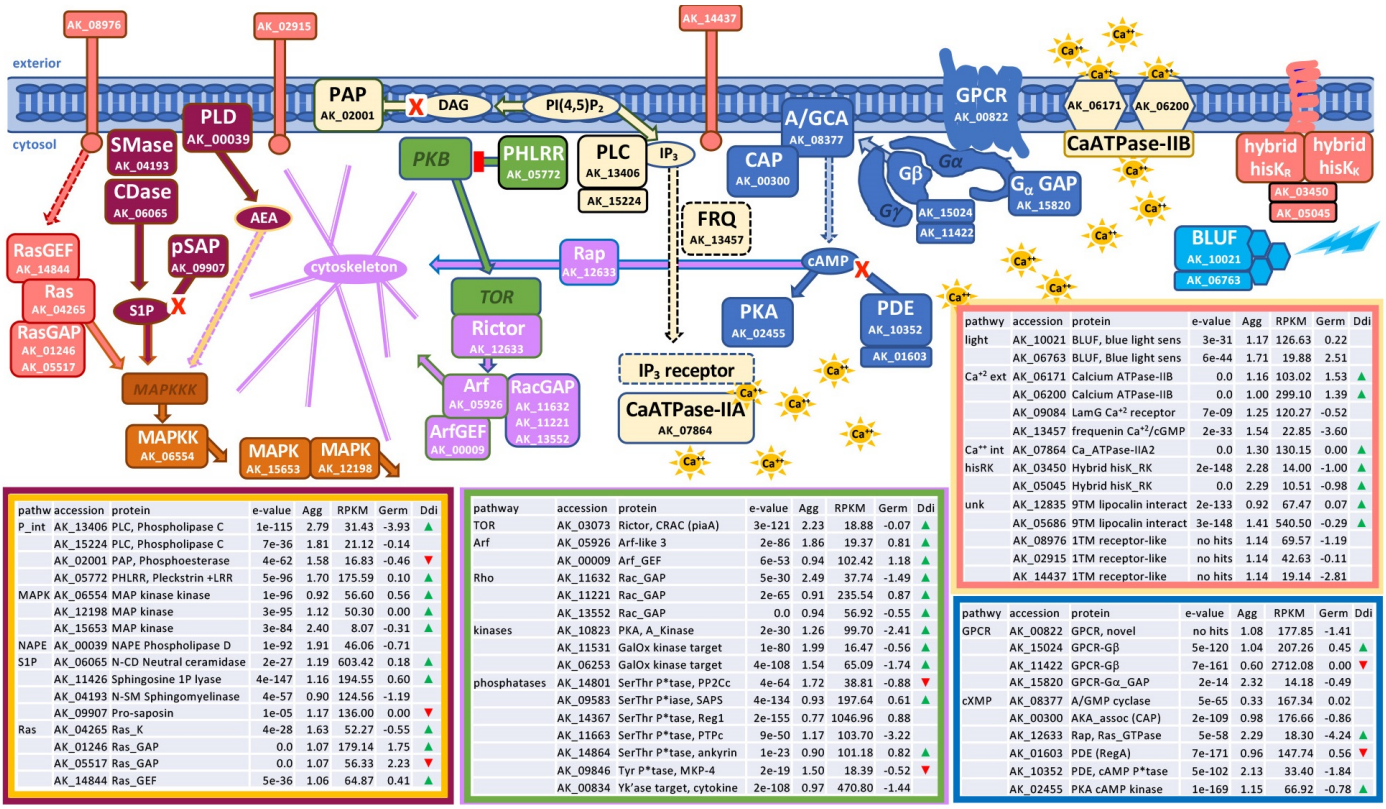


Figure 6

The *Acrasis kona* proteasome and its life-cycle specific expression compared to *Dictyostelium discoideum AX4* (Ddi AX4). Proteins were identified by BLASTp using human queries with hit strength indicated by color intensity (key at bottom right). Hits with e-values from e^{-25} to e^{-5} were scored only if they showed >50% query coverage (cvg). RNAseq length-corrected read numbers are shown for growing ($RPKM_{Gro}$) and aggregating ($RPKM_{Agg}$) *A. kona* and similar time points for Ddi AX4 ($RPKM_0$ and $RPKM_5$, respectively). Green and red arrows indicate increased ($DE_{Log2} \geq 0.9$) or decreased ($DE_{Log2} \leq -0.9$) gene expression for growth vs aggregation (DE_{Agg}) and aggregation vs germination (DE_{Germ}) in *A. kona* and for 0 vs 1 (DE_{0_1}) and 2 vs 5 (DE_{2_5}) hours of starvation in Ddi AX4 (time intervals with largest differences in expression). Accession numbers, BLASTp e-values and DE values are in Table S13. Taxon name

abbreviations: *A. kona* (Ako), *Aureococcus anophagefferens* (Aan), *Chlorella variabilis* (Cva), *D. discoideum AX4* (Ddi), *Naegleria gruberi* (Ngr), *Saccharomyces cerevisiae S288c* (Sce). DdiAX4 gene expression values were derived from Santhanam et al. (2015).



pathw	accession	protein	e-value	Agg	RPKM	Germ	Ddi
P_int	AK_13406	PLC, Phospholipase C	1e-115	2.79	31.43	-3.93	▲
	AK_15224	PLC, Phospholipase C	7e-36	1.81	21.12	-0.14	
	AK_02001	PAP, Phosphoesterase	4e-62	1.58	16.83	-0.46	▼
	AK_05772	PHLRR, Pleckstrin +LRR	5e-96	1.70	175.59	0.10	▲
MAPK	AK_06554	MAP kinase kinase	1e-96	0.92	56.60	0.56	▲
	AK_12198	MAP kinase	3e-95	1.12	50.30	0.00	▲
	AK_15653	MAP kinase	3e-84	2.40	8.07	-0.31	▲
NAPE	AK_00039	NAPE Phospholipase D	1e-92	1.91	46.06	-0.71	
S1P	AK_06065	N-CD Neutral ceramidase	2e-27	1.19	603.42	0.18	▲
	AK_11426	Sphingosine 1P lyase	4e-147	1.16	194.55	0.60	▲
	AK_04193	N-SM Sphingomyelinase	4e-57	0.90	124.56	-1.19	
	AK_09907	Pro-saposin	1e-05	1.17	136.00	0.00	▼
Ras	AK_04265	Ras_K	4e-28	1.63	52.27	-0.55	▲
	AK_01246	Ras_GAP	0.0	1.07	179.14	1.75	▲
	AK_05517	Ras_GAP	0.0	1.07	56.33	2.23	▼
	AK_14844	Ras_GEF	5e-36	1.06	64.87	0.41	▲

pathway	accession	protein	e-value	Agg	RPKM	Germ	Ddi
TOR	AK_03073	Rictor, CRAC (piaA)	3e-121	2.23	18.88	-0.07	▲
Arf	AK_05926	Arf-like 3	2e-86	1.86	19.37	0.81	▲
	AK_00009	Arf_GEF	6e-53	0.94	102.42	1.18	▲
Rho	AK_11632	Rac_GAP	5e-30	2.49	37.74	-1.49	▲
	AK_11221	Rac_GAP	2e-65	0.91	235.54	0.87	▲
	AK_13552	Rac_GAP	0.0	0.94	56.92	-0.55	▲
kinases	AK_10823	PKA, A_Kinase	2e-30	1.26	99.70	-2.41	▲
	AK_11531	GalOx kinase target	1e-80	1.99	16.47	-0.56	▲
	AK_06253	GalOx kinase target	4e-108	1.54	65.09	-1.74	▲
phosphatases	AK_14801	SerThr P*tase, PP2Cc	4e-64	1.72	38.81	-0.88	▼
	AK_09583	SerThr P*iase, SAPS	4e-134	0.93	197.64	0.61	▲
	AK_14367	SerThr P*tase, Reg1	2e-155	0.77	1046.96	0.88	
	AK_11663	SerThr P*tase, PTPc	9e-50	1.17	103.70	-3.22	
	AK_14864	SerThr P*tase, ankyrin	1e-23	0.90	101.18	0.82	▲
	AK_09846	Tyr P*tase, MKP-4	2e-19	1.50	18.39	-0.52	▼
	AK_00834	Yk'ase target, cytokine	2e-108	0.97	470.80	-1.44	

pathw	accession	protein	e-value	Agg	RPKM	Germ	Ddi
light	AK_10021	BLUF, blue light sens	3e-31	1.17	126.63	0.22	
	AK_06763	BLUF, Blue light sens	6e-44	1.71	19.88	2.51	
Ca^{+2} ext	AK_06171	Calcium ATPase-IIB	0.0	1.16	103.02	1.53	▲
	AK_06200	Calcium ATPase-IIB	0.0	1.00	299.10	1.39	▲
	AK_09084	LamG Ca^{+2} receptor	7e-09	1.25	120.27	-0.52	
	AK_13457	frequenin Ca^{+2}/cGMP	2e-33	1.54	22.85	-3.60	
Ca^{++} int	AK_07864	Ca_ATPase-IIA2	0.0	1.30	130.15	0.00	▲
hisRK	AK_03450	Hybrid hisK_RK	2e-148	2.28	14.00	-1.00	▲
	AK_05045	Hybrid hisK_RK	0.0	2.29	10.51	-0.98	▲
unk	AK_12835	9TM lipocalin interact	2e-133	0.92	67.47	0.07	▲
	AK_05686	9TM lipocalin interact	3e-148	1.41	540.50	-0.29	▲
	AK_08976	1TM receptor-like	no hits	1.14	69.57	-1.19	
	AK_02915	1TM receptor-like	no hits	1.14	42.63	-0.11	
	AK_14437	1TM receptor-like	no hits	1.14	19.14	-2.81	

pathw	accession	protein	e-value	Agg	RPKM	Germ	Ddi
GPCR	AK_00822	GPCR, novel	no hits	1.08	177.85	-1.41	
	AK_15024	GPCR-Gβ	5e-120	1.04	207.26	0.45	▲
	AK_11422	GPCR-Gβ	7e-161	0.60	2712.08	0.00	▼
	AK_15820	GPCR-Gα_GAP	2e-14	2.32	14.18	-0.49	
cXMP	AK_08377	A/GMP cyclase	5e-65	0.33	167.34	0.02	
	AK_00300	AKA_assoc (CAP)	2e-109	0.98	176.66	-0.86	
	AK_12633	Rap, Ras_GTPase	5e-58	2.29	18.30	-4.24	▲
	AK_01603	PDE (RegA)	7e-171	0.96	147.74	0.56	▼
	AK_10352	PDE, cAMP P*tase	5e-102	2.13	33.40	-1.84	
	AK_02455	PKA cAMP kinase	1e-169	1.15	66.92	-0.78	▲

Figure 7

Developmental signaling in *Acrasis kona.* *A. kona* accessions with substantially increased expression during aggregation (Aggup) are shown with accession numbers and grouped into common signaling pathways along with a few key enzymes with substantial but not substantially increased aggregation expression (RPKM >100). Signaling interactions are indicated as follows: canonical activation (solid arrows), predicted activation (dotted-line arrows), inhibition (red-blockhead arrows) and degradation (arrow capped with a red X). Second messenger molecules are shown in ovals as follows: cAMP (cyclic AMP), DAG (diacyl glycerol), IP_3 (inositol 3-phosphate), $PI(4,5)P_2$ (phosphoinositol 4,5-bisphosphate), AEA (anandamide), S1P (sphingosine 1-phosphate) and calcium (Ca^{++}). Tables are organized by general signaling pathways and show, from left to right, *A. kona* accession numbers, protein names, BLASTp e-values, DE_{Log2} for growth versus aggregation (Agg) and aggregation versus germination (Germ), RPKM values for aggregation (RPKM) and up- or down-regulation of Ddi AX4 homologs (green and red arrows, respectively). Protein names are abbreviated as follows: G-protein coupled receptor (GPCR), G-protein alpha (Gα), beta (Gβ), and gamma (Gγ) subunits, hybrid histidine kinase signal receptor (hybrid hisK), mitogen activated protein kinase (MAPK), N-acyl-phosphatidylethanolamine (PLD), small GTPases (Arf, Rac, Rap, Ras and Rho), GTPAse exchange factors (GEFs) and activators (GAPs), phosphatase (P*tase), tyrosine kinase (Yk'ase). Structure/function predictions for AK_00820 (GPCR) and AK_11986 (novel 1TM

receptor) are shown in Figure S9. Full details of individual sequence annotation and expression are shown in Table S17.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AkonaSupDatMovie1.mov
- AkonaSupDatTables01.xlsx
- AkonaSupDatFigs01.pdf
- AkonamsGraphicalAbstract.pdf