

To the Origin of Fungi. Analysis of MFS Transporters of First Assembled Aphelidium Genome Highlights Dissimilarity of Osmotrophic Abilities between Aphelida and Fungi

[Igor R. Pozdnyakov](#)*, Evgeniy V. Potapenko, [Elena S. Nassonova](#), [Vladislav V. Babenko](#), Daria I. Boldyreva, Victoria S. Tsvetkova, Sergey A. Karpov

Posted Date: 28 July 2023

doi: 10.20944/preprints202307.1928.v1

Keywords: Aphelida; fungi; holomycota; osmotrophy; MFS-proteins; evolution; genome; sugar porters



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

To the Origin of Fungi. Analysis of MFS Transporters of First Assembled *Aphelidium* Genome Highlights Dissimilarity of Osmotrophic Abilities between Aphelida and Fungi

Igor R. Pozdnyakov ^{1,*}, Evgeniy V. Potapenko ^{2,3}, Elena S. Nassonova ^{4,5}, Vladislav V. Babenko ⁶, Daria I. Boldyreva ⁶, Victoria S. Tsvetkova ⁵ and Sergey A. Karpov ¹

¹ Zoological Institute, Russian Academy of Sciences, St. Petersburg 199034, Russian Federation

² Institute of Evolution, University of Haifa, Haifa 3498838, Israel

³ Department of Evolutionary and Environmental Biology, University of Haifa, Haifa 3498838, Israel

⁴ Laboratory of Cytology of Unicellular Organisms, Institute of Cytology, Russian Academy of Sciences, St. Petersburg 194064, Russian Federation

⁵ Department of Invertebrate Zoology, Faculty of Biology, St. Petersburg University, St. Petersburg 199034, Russian Federation

⁶ Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine, Federal Medical Biological Agency, Moscow, Russian Federation

* Correspondence: d_igor_po@yahoo.com; Tel.: +79062469262

Abstract: Aphelids are a holomycotan group, represented exclusively by parasitoids infecting algae. They form a sister lineage to Fungi in the phylogenetic tree and represent a key group for reconstruction of the evolution of Holomycota and for analysis of the origin of Fungi. The newly assembled genome of *Aphelidium insullamusi* (Holomycota, Aphelida) with a total length of 18.9Mb, 7820 protein-coding genes and a GC percentage of 52.05% was obtained by a hybrid assembly based on Oxford Nanopore long reads and Illumina paired reads. In order to trace the origin and the evolution of fungal osmotrophy and its presence or absence in Aphelida, we analyzed the set of main fungal transmembrane transporters, which are proteins of the Major Facilitator superfamily (MFS), in the predicted aphelid proteomes. This search has shown an absence of a specific fungal protein family Drug:H⁺ antiporters-2 (DAH-2) and specific fungal orthologs of the sugar porters (SP) family, and the presence of common opisthokont's orthologs of the SP family in four aphelid genomes. The repertoire of SP orthologs in aphelids turned out to be less diverse than in free-living opisthokonts, and one of the most limited among opisthokonts. We argue that aphelids do not show signs of similarity with fungi in terms of their osmotrophic abilities, despite the sister relationships of these groups. Moreover, the osmotrophic abilities of aphelids appear to be reduced in comparison with free-living unicellular opisthokonts. Therefore, we assume that the evolution of fungi-specific traits began after the separation of fungal and aphelid lineages, and there are no essential reasons to consider aphelids as a prototype of the fungal ancestor.

Keywords: Aphelida; fungi; holomycota; osmotrophy; MFS-proteins; evolution; genome; sugar porters

1. Introduction

Aphelida is a group of algal parasitoids represented by an intracellular ameboid-plasmodial trophic stage and exiting to the environment zoospores that infect other algal cells [1–3]. Phylum Aphelida belongs to Holomycota, one of two branches of Opisthokonta [4,5]. and, as shown by phylogenomic analysis, is a sister group to Fungi, very distinctive and specialized osmotrophic organisms [6–9].

After the exclusion of the fungi-like stramenopiles, Fungi became a distinct monophyletic taxon within Opisthokonta, uniting osmotrophic organisms with mycelial or pseudomycelial organization and a chitinous cell wall [4,10–12]. Unambiguous characterization of this taxon is difficult because of secondary changes in some of its representatives [13]. Conventionally, "higher", or crown fungi are completely devoid of flagella, while the "lower", or basal fungi have a flagellated stage in their life cycle, the zoospores [4,10]. The sister relationship of Aphelida and Fungi seems surprising because

the representatives of the former group, being amoeboid and phagotrophic, have only superficial similarities in the life cycle with the members of the latter one [2]. Meanwhile, aphelids are closer to fungi than to *Rozella* spp.; which are similar to algal parasitoids both in life cycle, cell morphology, and phagotrophic mode of feeding [5,6,9].

The fact that fungi are related to such different organisms gives a possibility to identify early stages of fungal specialization since it remains unclear how the fungal features originated and evolved. Obviously, the fungal traits could not arise simultaneously, but had to increase gradually until they reached a pronounced morphological appearance. Therefore, it is possible that the organisms related to fungi could retain some fungal features in an ancient, inchoate, or not fully formed state, having inherited them from a common ancestor with fungi.

In the field of taxonomy, such searches and finds can clarify the frames of the Fungi, since the problem of apomorphies and borders of this taxon has not yet been resolved. Some "classic" fungal signs, like osmotrophy, are not only characteristic of Fungi. Other features, such as mycelial growth and chitin cell wall, are repeatedly lost in undoubted representatives of fungal lineages [13]. The idea of fungal apomorphies became even more uncertain after the inclusion of phagotrophic aphelids, *Rozella*, and microsporidia in Fungi [7]. The most recent and comprehensive review of eukaryotic taxonomy states: "There are no unambiguous morphological, subcellular, or biochemical synapomorphies of fungi." [4]. The detection or non-detection of any latent features of fungi in closely related organisms may be an additional argument pro or contra their inclusion in the Fungi.

We suggested that aphelids can have increased osmotrophic capabilities compared to free-living opisthokonts, despite the fact that they retain phagotrophic nutrition. We had several reasons for this assumption. (1) A common ancestor of aphelids and fungi could already have a heightened capacity for osmotrophy, which could have been inherited by the aphelids. (2) For aphelids, the increase in osmotrophy may be adaptive, since their trophonts are immersed in the host cytoplasm, which is rich in nutrient molecules. (3) The repertoire of proteins associated with digestive vacuoles, especially the COMMD/CCDC22/CCDC93 (CCC) complex, is somewhat reduced in aphelids [9]. This reduction may be a synapomorphy of both, the aphelids and fungi, inherited from a common ancestor. Such a reduction can be justified precisely in the case when the osmotrophic mechanism of nutrition was strengthened in the common ancestor. To verify this assumption, we searched for Major Facilitator Superfamily (MFS) proteins, which are a key component of osmotrophic machinery, in the predicted aphelid proteomes, and carried out their comparative analysis.

The MFS superfamily is a large group of plasma membrane proteins that are present in all cellular organisms and serve as transmembrane transporters of various substances, including mono- and oligosaccharides, metabolites, amino acids, and oxyanions [14,15]. The main feature of MFS transporters is the presence of usually 12 (sometimes 14) transmembrane helices, which are divided into two 6-helix blocks opposite each other and connected by an extended loop. This molecular composition is often referred to as the MFS domain [15–17]. The MFS superfamily is divided into different families, whose proteins differ in their mechanism of operation, transported substrates, and features of function [15,16,18].

MFS transporters in fungi are numerous and diverse because fungi, as obligate osmotrophic organisms, must transport all types of nutrient substrates across the membrane [19–21]. MFS proteins in fungi are represented both by families common to all organisms, such as Sugar Porters (SP), and by fungal-specific families, such as Drug:H⁺ Antiporters-2 (DHA-2) [18,19,22–24]. Multiple and specialized SP proteins perform a function of transport of various carbohydrate substrates. DHA-2 proteins carry various substances, including nitrogen-containing ones. Thus, the proteins of these two MSF families play a major role in fungal nutrition.

In this article, we present de novo assembled genome of *Aphelidium insulamus*. Based on the predicted proteomes derived from the available aphelid genomes and transcriptomes, we searched for SP and DHA-2 proteins and compared them qualitatively and quantitatively with corresponding proteins of dikaryan and zoosporic fungi, the unicellular parasitoid holomycotan *Rozella allomycis*, and representatives of Holozoa.

2. Materials and Methods

2.1. DNA extraction and genome sequencing

The DNA was extracted from the strain X-133 of *Aphelidium insulamus* maintained in the culture collection of parasitic protists (CCPP) of Zoological Institute Russian Academy of Sciences (ZIN RAS) [25] using two different protocols. In the frames of the first protocol, DNA was purified from the heavily infected culture of *Tribonema gayanum*, containing almost digested algal cells with trophonts and plasmodia of *Aph. insulamus*. DNA was extracted using Wizard® Genomic DNA Purification Kit (Promega, USA) according to the manufacturer's protocol. Purified DNA was used for genome sequencing on the Oxford Nanopore platform.

According to the second protocol, DNA was extracted from zoospores, which were isolated from the infected algal culture and concentrated by centrifugation at Eppendorf MicroSpin centrifuge at 4300 g for 10 minutes. Sedimented cells were used for the Multiple Displacement Amplification (MDA) with Repli-g Single Cell Amplification Kit (Qiagen, Germany), according to the manufacturer's protocol for single-cell reactions. To avoid biases of uneven whole genome amplification, ten independent MDA reactions were performed, each yielded ca 5.5-8.0 µg DNA. The MDA products were checked for the presence of target DNA using PCR amplification of a fragment of the SSU rRNA gene with universal eukaryotic primers S12.2 and RibB [26]. PCR amplification program: 5 min denaturation at 94 °C, 35 cycles of a denaturation step at 94 °C for 15 s, a 30 s annealing step at 50 °C and an extension step at 72 °C for 2 min, and a final elongation step of 7 min at 72 °C. The positive DNA samples were mixed in equimolar proportion and used for library preparations and sequencing on Illumina HiSeq4000 and Oxford Nanopore platforms.

For Illumina HiSeq4000 sequencing two paired-end libraries were prepared following the TruSeq and Nextera library preparation protocols with an insert length of 700 bp. A total of 62 million and 83 million paired-end reads were obtained for the two libraries.

The long reads were generated with MinION and PromethION sequencing (Oxford Nanopore Technologies, UK). The sequencing libraries were prepared using the ligation sequencing kit SQK-LSK109, native barcoding expansion kit EXP-NBD104 and EXP-NBD114. The gained library for ONT sequencing was then loaded into the flow cells (FLO-MIN106 and FLO-PRO002). Thus, two libraries were obtained with 560 000 and 300 000 long reads.

2.2. Genome assembly and annotation

The initial genomic assemblies were performed with Flye, v.2.9.1 [27] with default settings using two libraries. The draft assembly was checked for contaminations with BlobToolKit v2.3.3 [28]. Trusted contigs were selected based on the annotation of contigs against NCBI nucleotide and UniProt reference proteomes databases, GC content, and short/long reads coverage information. Further, both long reads libraries were mapped on the trusted contigs using the Minimap 2, v.2.24 [29] and all unmapped reads were discarded. The next step assembly was made with the trusted reads with Flye, v.2.9.1 using the same way. The new assembly was polished using the Illumina paired read libraries. The first step of polishing was made in Racon, v.1.4.3 [30], and the second in Pilon, v.1.24 [31]. The assembly quality was controlled with BUSCO, v.5.4.2 [32] and QUAST, v.5.0.2 [33]. on every step of assembly.

Structure annotation of assembled genome was carried out using funannotate pipeline, v.1.8.13 [34], which includes repeat masking with tatan, ab initio gene-prediction training (Augustus, PASA, SNAP, GlimmerHMM, GeneMark), generating consensus gene model (Evidence Modeler) [35] and functional annotating of proteins against several databases (Pfam, InterPro, GO, dbCAN, BUSCO, MEROPS, EggNog, COG). We improved gene prediction with protein evidence from UniProt database and transcript evidence datasets of closely related species: *Aphelidium insulamus* X-134_O14 [25] and *Aphelidium tribonematis* X-102_P2.

Secreted proteins were estimated with Phobius web service [36] and tRNAs were predicted in silico with tRNAscan-SE algorithm [37] included in the funannotate pipeline. Structure and functional comparison with genomes of related species (*Amoebophilum protococcarum*,

Amoebophilidium occidentale) and zoosporic fungi (*Gonapodya prolifera*, *Blyttomyces helicus*, *Powellomyces hirtus*, *Spizellomyces punctatus*) were performed with compare command of funannotate pipeline. All considered genomes before analysis were re-annotated with InterProScan [38] on the same databases versions.

For functional comparison, all analyzed genomes were (re)annotated with InterProScan. To estimate PFAM motifs occurrence in holomycotan genomes we applied non-metric multidimensional scaling (NMDS) projection of a Bray-Curtis distance matrix implemented in compare command of funannotate pipeline.

2.3. The selection and analysis of MFS-domain proteins

The MFS-domain proteins were initially chosen by the BLAST searching in the predicted proteomes of selected organisms with annotated genomes. For the search, the web interface of NCBI BLASTP [39] was used. The proteins of SP family were searched in the predicted proteomes of genomes (NCBI genome ID in brackets) of dikaryan fungi *Saccharomyces cerevisiae* (559292), *Neurospora crassa* (367110), *Cryptococcus neoformans* var. *neoformans* (214684), zoosporic fungi *Gonapodya prolifera* (1344416), *Spizellomyces punctatus* (645134), the sponge *Amphimedon queenslandica* (2698) and the knidarian *Stylophora pistillata* (18227). The *S. cerevisiae* glucose sensor NP_010087.1 was taken as the initial query. The proteins of DAH-2 family were searched in the aforementioned fungal genomes with the initial query NP_011740.3 which is *S. cerevisiae* azole transporter. The E-value and bit-scores threshold were set to 1e-05 and 50, correspondently.

Among the BLASTP hits the proteins of desired families were selected based on annotations and the found sequences were downloaded as two databases for SP proteins and DAH-2 proteins separately.

The hidden Markov models were built by the hmmbuild program of the hmmer, v.3.3.2 batch [40] for SP and DAH-2 proteins separately. The test search using hmmsearch program against the proteins of the used proteomes and the additionally predicted proteomes of *Agaricus bisporus* var. *bisporus* (936046), *Blyttomyces helicus* (388810), *Powellomyces hirtus* (109895), *Salpingoeca rosetta* (946362), *Arabidopsis thaliana* (3702), *Escherichia coli* (167), *Acetivibrio nasutus* (1720083) shows that bitscore threshold 250 allows clear find the proteins of desired families and cut off the proteins of the other families, even related ones.

At the next step the SP and DAH-2 proteins were chosen from the predicted proteomes of *Amoebophilidium protococcarum* (114058), *Amoebophilidium occidentale* (114059), *Paraphelidium tribonematis* and *Aphelidium insulamus*, as well as the predicted proteomes of *Agaricus bisporus* var. *bisporus* (936046), *Blyttomyces helicus* (388810), *Powellomyces hirtus* (109895), *Synchytrium endobioticum* (286115), *Salpingoeca rosetta* (946362), *Monosiga brevicollis* (431895), *Capsaspora owczarzaki* (595528) and *Rozella allomyces* (12422) by the searching of correspondent profiles against these proteomes using hmmsearch. The predicted proteome of *P. tribonematis* is absent in the NCBI databases and it was downloaded from the Figshare store [6]. The predicted proteome of *Aph. insulamus* was obtained using the genome sequenced and assembled in the frames of the present study and the pipeline described above.

The phylogenetic analysis was performed for the found sequences of sugar transporters. The multiple sequence alignment (MSA) was prepared in the M-Coffee aligner using the web server interface (<https://tcoffee.org>) [41]. The MSA was treated in TrimAl, v.1.4.rev15 [42] with a gap threshold of 0.5 for the removal of columns with the gap abundances appeared due to the large divergence between protein sequences.

The initial tree was constructed on the ground of trimmed MSA using IQ-Tree 2, v.2.0.3 [43] with the settings of automated determination of the substitution model and 1000 replicates of ultrafast bootstrap. The most suitable substitution model was determined as LG+F+G4. For the final tree construction was used IQ-Tree 2, v.2.0.3 also, with LG+F+G4 substitution model and 100000 replicates of ultrafast bootstrap.

The sequences of bacterial SP proteins (*Bacteroides* bacterium, *Acetivibrio nasutus*, *Bacteroides* bacterium, *Lentivibrio nasutus* spp., *Secundivibrio nasutus* hailunensis) showing the best

BLAST matches for some aphelid proteins were downloaded from the ncbi database by their accessions. A phylogenetic tree including aphelid, fungal and bacterial SP proteins was constructed as described above.

Prediction of transmembrane structures in SP proteins was performed using web service TMHMM-2.0 (<https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>).

Since in case of *Am. protococcarum* the search showed the presence of 8 separate SP proteins, which form two quartets on the phylogenetic tree, we specifically assessed the level of similarity of these proteins using reciprocal BLASTP alignments.

3. Results

3.1. Assembly of *Aphelidium insulamus* genome

The hybrid assembly based on Oxford Nanopore and Illumina paired reads yielded an *Aph. insulamus* genome with a total length of 18,927,283 bp that were distributed among 274 scaffolds, with an N50 252,907 and a 52.05% GC percentage (NCBI BioProject accession number PRJNA902644). The assembled genome contains 5% of repetitive sequences. The average coverage of Nanopore long reads and Illumina paired reads were about 9,6-fold and 95,2-fold, respectively. We identified 7925 genes, including 7820 protein-coding ones. The average lengths of the predicted genes and proteins were 1664 nucleotides and 484 amino acids, respectively. Funannotate pipeline annotated the following number of predicted protein-coding sequences using different databases: InterPro (5893), Pfam (4957), GO (4439), EggNOG (3735), BUSCO Eukaryota Odb10 (287), MEROPS (268) and dbCAN (125). In addition, we annotated 1431 secreted proteins and 105 tRNA-encoding genes.

The comparison of obtained assembly with the genome assemblies of two related aphelids and a selected set of zoosporic fungi in terms of key indicators showed its comparable characteristics (Table 1). The quality of assembly of *Aph. insulamus* is significantly inferior to only two assemblies: *Amoeboaphelidium protococcarum* and *Spizellomyces punctatus*. The quantitative indicators (assembly size, scaffolds number, N50, average and largest scaffolds, unique BUSCOs) of *Aph. insulamus* assembly are between the corresponding indicators of previously published genomes of *Am. protococcarum* and *Am. occidentale* [9].

Table 1. Comparison of the genome assemblies of *Aphelidium insulamus* and closely related species.

| Species | Assembly size (bp) | Num scaffolds | Scaffold N50 (bp) | Average scaffold (bp) | Largest scaffold (bp) | GC, % | Num genes | Num proteins | Unique BUSCOs |
|---------------------------------------|--------------------|---------------|-------------------|-----------------------|-----------------------|-------|-----------|--------------|---------------|
| <i>Aphelidium insulamus</i> | 18,927,283 | 274 | 252,907 | 69,078 | 1,020,338 | 52.05 | 7,925 | 7,820 | 277 |
| <i>Amoeboaphelidium protococcarum</i> | 24,734,778 | 258 | 2,170,272 | 95,871 | 3,250,117 | 40.50 | 13,180 | 13,180 | 281 |
| <i>Amoeboaphelidium occidentale</i> | 13,559,732 | 951 | 73,507 | 14,258 | 366,412 | 39.93 | 7,568 | 7,495 | 277 |
| <i>Gonapodya prolifera</i> | 48,794,828 | 352 | 347,324 | 138,622 | 1,572,201 | 51.75 | 13,911 | 13,831 | 283 |
| <i>Blyttomyces helicus</i> | 46,468,912 | 8,398 | 6,675 | 5,533 | 73,981 | 53.75 | 12,446 | 12,167 | 168 |
| <i>Powellomyces hirtus</i> | 26,238,698 | 482 | 157,542 | 54,437 | 764,225 | 51.37 | 6,536 | 6,536 | 292 |
| <i>Spizellomyces punctatus</i> | 24,131,112 | 38 | 1,465,700 | 635,029 | 2,242,449 | 47.16 | 9,164 | 9,422 | 295 |

Interestingly, GC content of *Aph. insulamus* genome is more similar to fungi than to *Amoeboaphelidium* species. On the opposite, the number of proteins and BUSCO genes is closer to *Aph. occidentale*.

To arrange species in two-dimensional space based on the functionality of their genomes, we applied the Nonmetric Multidimensional Scaling (NMDS) algorithm to functional domain annotations, using abundance data of PFAM domains (Figure 1). We identified 3 main clusters. *Gonapodya prolifera*, *Spizellomyces punctatus* and *Powellomyces hirtus* form the fungi cluster, then

Amoeboaphelidium occidentale, *Aphelidium insulamus* and *Amoeboaphelidium protococcarum* form the aphelid cluster, and the third one contains only *Blyttomyces helicus*. We assume that the lack of clusterization of *B. helicus* with the rest of fungi is explained by the poor quality of its assembly. Results showed that *Aph. insulamus* is very closed to *Am. occidentale* in the functional domains content while *Am. protococcarum* is rather distant from both of them, which may be due to the established genome-wide duplication of *Am. protococcarum* [9].

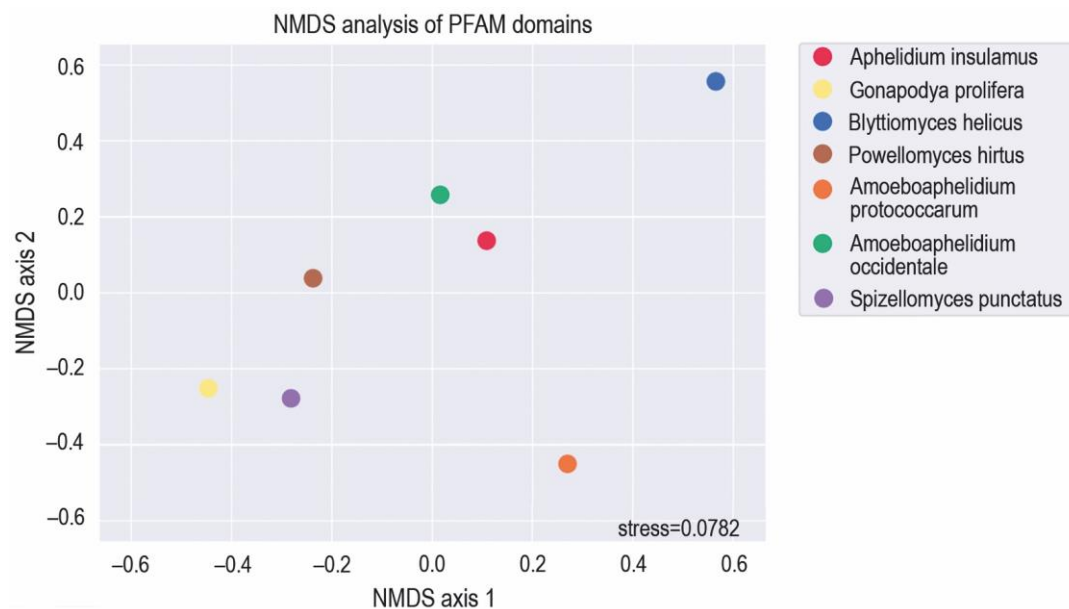


Figure 1. NMDS analysis showing PFAM domains co-occurrence in the genomes of *Aphelidium insulamus*, related aphelid species and zoosporic fungi.

3.2. MFS Protein Analysis

Proteins of the DHA-2 family have not been found in any species of Aphelida, as well as in *R. allomyces*. Search for SP family proteins in aphelid genomes found only one SP protein in *P. tribonematis*, two SP proteins in *Am. occidentale*, three proteins in *Aph. insullamus*, and eight ones in *Am. protococcarum* with a duplicated genome (see Discussion). The detected eight SP-porters of *Am. protococcarum* are divided into two quartets of proteins with a very high level of identity (93-98%) within each, which shows their recent origin from two genes through a series of duplications.

The number of SP proteins comparable to that in aphelids was found in *C. owczarzewski* (1), *R. allomyces* (2), and *B. helices* (3). In all other studied opisthokonts their number is higher (the case with *Am. protococcarum* will be considered separately), e.g., Metazoa have more than a dozen of them and dikaryan fungi contain several dozens of SP proteins (Table 2).

Table 2. The total numbers and accession numbers (NCBI) of SP proteins found in studied opisthokont species. Large groups are marked with colors.

| | | | | | | |
|----------|-----------------|----------------|-----------|------------|------------------|---------|
| | | | | | | |
| Aphelida | Zoosporic fungi | Dikaryan fungi | Rozellida | Filasterea | Choanoflagellata | Metazoa |

| Species | Designation in the phylogenetic tree | Number of SP proteins | Accession numbers |
|---------------------------------------|--------------------------------------|-----------------------|---|
| <i>Aphelidium insulamus</i> | aphins | 3 | OR609_004515-T1 (FUN_002943-T1), OR609_000014-T1 (FUN_001752-T1), OR609_006691-T1 (FUN_007167-T1) |
| <i>Paraphelidium tribonematis</i> | partrib | 1 | Partr_v1_DN257 |
| <i>Amoeboaphelidium protococcarum</i> | amoebprot | 8 | KAI3642224.1, KAI3631285.1, KAI3647471.1, KAI3641405.1, KAI3636887.1, KAI3651751.1, KAI3653019.1, KAI3638028.1 |
| <i>Amoeboaphelidium occidentale</i> | amoeboccid | 2 | KAI3658561.1, KAI3659551.1 |
| <i>Blyttiomycetes. helicus</i> | blitt | 3 | RKO83577.1, RKO87217.1, RKO84726.1 |
| <i>Gonapodya prolifera</i> | gonap | 6 | KXS20085.1, KXS21772.1, KXS18310.1, KXS19020.1, KXS14312.1, KXS18846.1 |
| <i>Powellomyces hirtus</i> | powell | 4 | TPX57964, TPX61303, TPX53442, TPX57119 |
| <i>Spizellomyces punctatus</i> | spizel | 5 | XP_016610136.1, XP_016608866.1, XP_016607884.1, XP_016608771.1, XP_016607634.1 |
| <i>Synchytrium endobioticum</i> | synchyt | 4 | TPX52967.1, TPX40513.1, TPX48127.1, TPX35163.1 |
| <i>Agaricus bisporus</i> | agar | 26 | XP_006454719.1, XP_006461850.1, XP_006453951.1, XP_006456284.1, XP_006459401.1, XP_006454620.1, XP_006457533.1, XP_006460576.1, XP_006459994.1, XP_006459113.1, XP_006463002.1, XP_006459487.1, XP_006459489.1, XP_006459491.1, XP_006458691.1, XP_006459490.1, XP_006456584.1, XP_006460947.1, XP_006455907.1, XP_006455078.1, XP_006456754.1, XP_006456796.1, XP_006456892.1, XP_006463354.1, XP_006459525.1, XP_006458285.1 |
| <i>Neurospora crassa</i> | neurosp | 34 | XP_959573.2, XP_959411.2, XP_965713.1, XP_955977.1, XP_962392.2, XP_958069.2, XP_960171.3, XP_956001.1, XP_959563.1, XP_960000.1, XP_964364.3, XP_958937.1, XP_956491.3, XP_959616.1, XP_959582.3, XP_961039.2, XP_963898.1, XP_963341.3, XP_963320.3, XP_960136.3, XP_960988.3, XP_961779.1, XP_960559.2, XP_958139.1, XP_955927.2, XP_964352.1, XP_963873.1, XP_965499.1, XP_960547.2, XP_963122.1, XP_957424.1, XP_001728155.2, XP_963801.1, XP_957394.1 |
| <i>Saccharomyces cerevisiae</i> | sacch | 30 | NP_010087.1, NP_010143.1, NP_011960.2, NP_116644.1, NP_013724.1, NP_013182.1, NP_011962.1, NP_010632.1, NP_012316.1, NP_010629.3, NP_010630.1, NP_011964.1, NP_014486.1, NP_012321.1, NP_014470.1, NP_010845.1, NP_010036.1, NP_012692.3, NP_014081.1, NP_010825.3, NP_010785.1, NP_014538.2, NP_012694.1, NP_010034.1, NP_011805.3, NP_116613.1, NP_009857.1, NP_010675.3, NP_009800.1, NP_011411.3 |

| Species | Designation in the phylogenetic tree | Number of SP proteins | Accession numbers |
|---------------------------------|---|-----------------------------|--|
| <i>Rozella allomycis</i> | roz | 2 | EPZ36062, EPZ31040 |
| <i>Capsaspora otczarzaki</i> | capsa | 1 | XP_004342757.1 |
| <i>Monosiga brevicollis</i> | monos | 5 | XP_001748408.1, XP_001745016.1, XP_001745420.1, XP_001748510.1, XP_001744797.1 |
| <i>Salpingoeca rosetta</i> | salpin | 5 | XP_004992344.1, XP_004991932.1, XP_004990085.1, XP_004994833.1, XP_004991367.1 |
| <i>Amphimedon queenslandica</i> | amphim | 13 | XP_019857806.1, XP_019857805.1, XP_019851521.1, XP_011406421.1, XP_003389392.2, XP_003383220.1, XP_019863478.1, XP_011402949.1, XP_019856859.1, XP_003384062.3, XP_003385376.1, XP_011408594.2, XP_003384855.1 |

Prediction of the structures of aphelid SP proteins (Figure 2) shows that all of them have a typical structure with a canonical MSF domain with 12 transmembrane helices. This shows that the aphelid SP proteins have the correct structure for this protein family and, most likely, function normally.

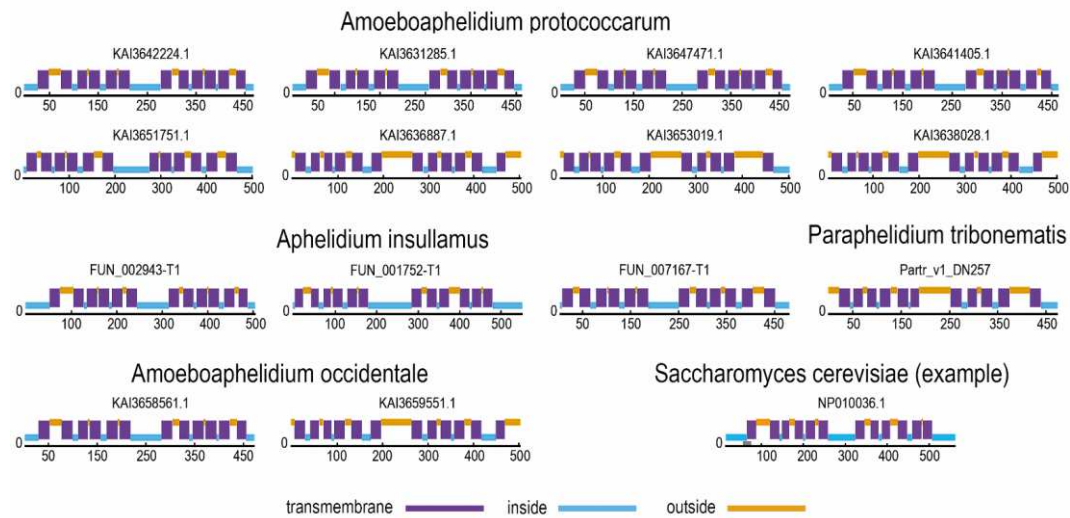


Figure 2. Prediction of distribution of intracellular, extracellular and transmembrane regions in the SP-protein molecule. The probable nature of the region is indicated by color (blue – intracellular, violet – transmembrane, orange – extracellular).

In the unrooted phylogenetic tree inferred from the analysis of aligned sequences of SP proteins, four variants of clades were observed: fungal, holozoan, aphelid, and mixed; the latter ones include the orthologous proteins of opisthokonts from different groups (Figure 3; short version, the full version in Newick format is available in the Supplementary materials: file Tree1.txt). Three holozoan clades (Figure 3) include proteins of either choanoflagellates or choanoflagellates and the sponge *A. queenslandica*. Two mixed clades (Figure 3) include proteins of all the studied species: fungi, aphelids, *R. allomycis*, and all holozoans. One (rarely two) proteins of each organism are present in these clades, except for *Am. protococcarum*. The branching order observed within the mixed clades is consistent with recent multigene phylogenies [8,9]. Specifically, the aphelid proteins form a sister branch to the fungal proteins, while the proteins of *R. allomycis* represent a sister lineage to the group uniting proteins of fungi and aphelids. In addition, the holozoan protein group is observed as a sister lineage

to the group formed by holomycotan proteins. In holozoan part of the clades the proteins of *C. owczarzaki* branched off before the Choanoflagellata and Metazoa lineages (Figure 4). Three observed fungal clades (Figure 3) have many branching levels and contain many or at least several proteins of each fungal species grouped into internal clusters. Within them, specific subbranches of dikaryan fungi, zoosporic fungi, and branches with proteins of both groups can be distinguished. The only aphelid clade (Figure 3) contains proteins of all aphelid species and does not contain proteins of other opisthokonts.

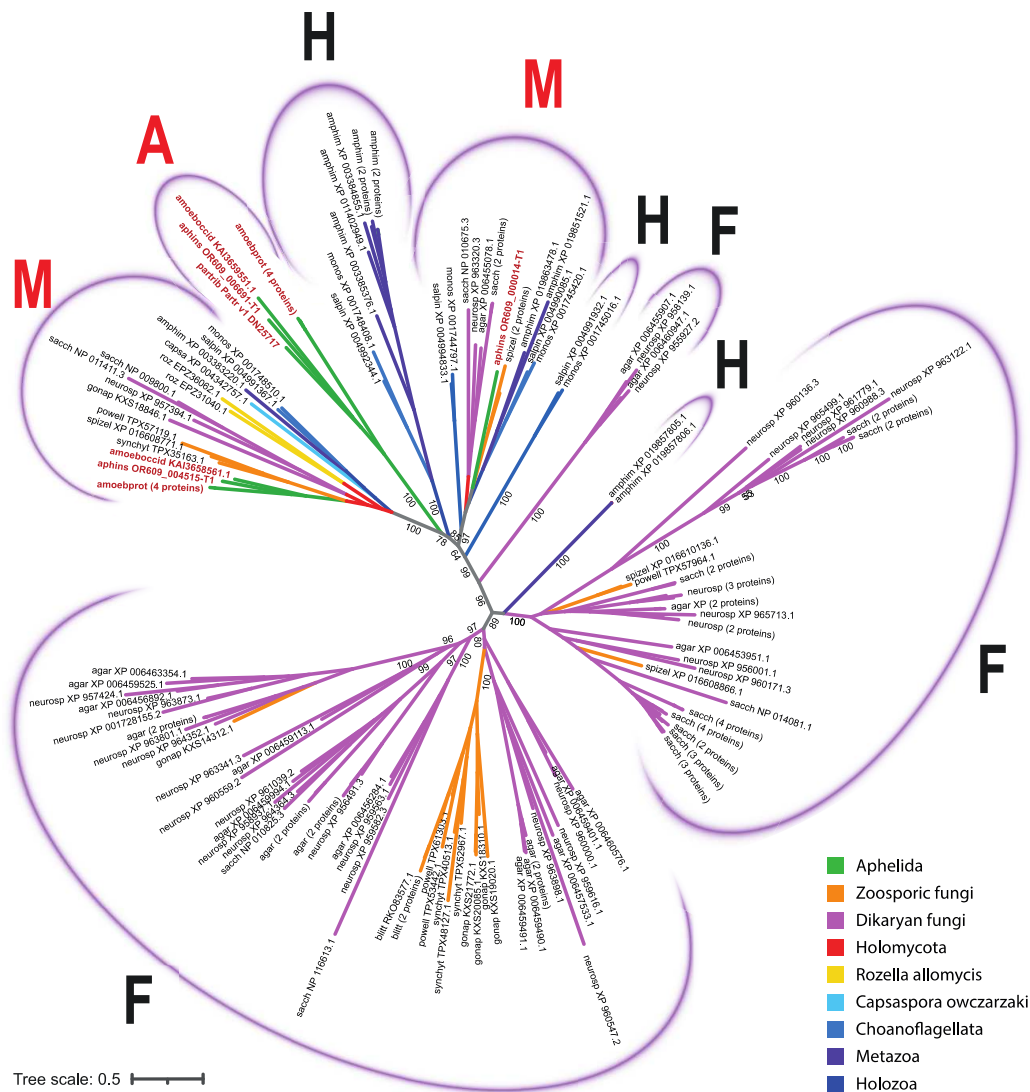


Figure 3. The phylogenetic tree of SP proteins of the studied opisthokont species. The opisthokont taxa are marked with color. The label of leaf is the short designation of species (see Table 2) and the accession number of the sequence in NCBI. Aphelid sequences are labelled in red. The numbers indicate the support values (100000 replicates of ultrafast bootstrap). A – aphelid clades, containing aphelid sequences only; F – fungal clades, containing sequences of zoosporic and dikaryan fungi only, H – holozoan clades, containing sequences of holozoan representatives; M – mixed clades, containing sequences of studied organisms both from Holomycota and from Holozoa.

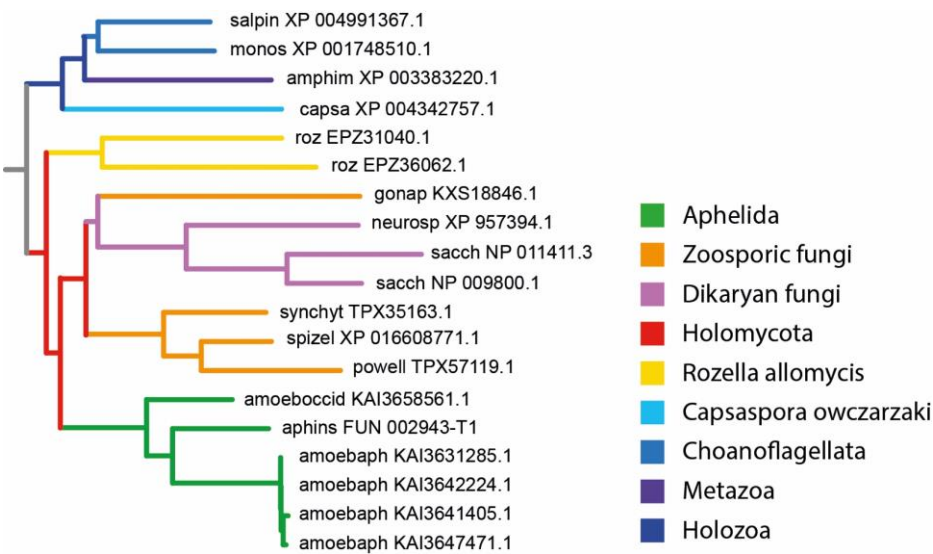


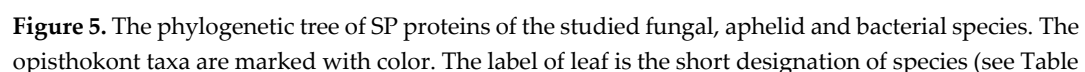
Figure 4. The branching order in the tree inferred from the analysis of orthogroup of SP-proteins identified in the studied opisthokont species. Species designations used in the phylogenetic tree are provided in Table 2.

Thus, aphelid protein sequences fall into three clades: a specific one with aphelid sequences only (A), and two mixed ones with holomycotan and holozoan sequences (M). At the same time, the clades combining the protein sequences of aphelids and fungi without any holozoan sequences are absent.

Protein sequences from the aphelid-specific clade show maximum similarity with SP sequences of bacteria and two proteins of dictyostelid amoebae in the BLAST search (Table 3), but do not show similarities with SP sequences of other opisthokonts which could indicate the horizontal transfer of these genes from bacteria to aphelids and some amoebas. However, in the phylogenetic tree of SP proteins including fungi, aphelids, and SP proteins of bacteria showing the best match in BLAST, the proteins of the aphelid-specific clade group neither with bacterial nor with fungal proteins (Figure 5; short version, the full version in Newick format is available in the Supplementary materials: file Tree2.txt). This observation suggests a complex and not yet understood evolutionary history of these proteins.

Table 3. Fifteen first hits of BLAST search for, A. protococcarum protein KAI3651751.1 as a query in general nucleotide database (nt). Blue letters – Aphelida, green letters – Amoebozoa, russet letters -- Bacteria.

| Description | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|-----------|-------------|-------------|---------|------------|----------|---------------------|
| <u>hypothetical protein MP228_003054</u> <u>[Amoeboaphelidium protococcarum]</u> | 1009 | 1009 | 100% | 0.0 | 100.00% | 501 | <u>KAI3651751.1</u> |
| <u>hypothetical protein MIR68_003639</u> <u>[Amoeboaphelidium protococcarum]</u> | 882 | 882 | 100% | 0.0 | 92.66% | 503 | <u>KAI3638028.1</u> |
| <u>hypothetical protein MP228_002444</u> <u>[Amoeboaphelidium protococcarum]</u> | 881 | 881 | 100% | 0.0 | 92.64% | 502 | <u>KAI3653019.1</u> |
| <u>hypothetical protein MIR68_005154</u> <u>[Amoeboaphelidium protococcarum]</u> | 863 | 863 | 100% | 0.0 | 90.89% | 504 | <u>KAI3636887.1</u> |
| <u>hypothetical protein MP638_005237</u> <u>[Amoeboaphelidium occidentale]</u> | 196 | 196 | 89% | 2e-52 | 30.95% | 483 | <u>KAI3659551.1</u> |
| <u>sugar transporter family protein</u> <u>[Tieghemostelium lacteum]</u> | 197 | 197 | 93% | 7e-52 | 30.42% | 631 | <u>KYO90579.1</u> |



2) and the accession number of the sequence in NCBI. The numbers indicate the support values (100000 replicates of ultrafast bootstrap).

4. Discussion

Up to now the genomes of three strains of aphelids have been sequenced and annotated: *Am. protococcarum* strains X5 and FD95, and *Am. occidentale* [9]. The first assembled genome of a representative of the genus *Aphelidium* provided new data, which we used for a comparative study of MFS proteins in Opisthokonta.

The species *Aph. insulamus* and *Am. occidentale* exhibit a high degree of similarity, not only in terms of the core gene set and the number of protein-coding genes but also in terms of functionality, as evidenced by the similarity of their domain profiles. *Am. protococcarum* is different from these two species. This observation correlates with a long distance between *Am. occidentale* and *Am. protococcarum* in the phylogenetic trees and with demonstrated polyphyly of the genus *Amoeboaphelidium* [9].

Aphelid genomes contain SP protein genes belonging to three different orthogroups. *Aph. insulamus* contains the genes of all three orthogroups, showing maximum diversity. *Am. protococcarum* and *Am. occidentale* contains genes belonging to two different orthogroups. In *P. tribonematis*, only one sequence encoding the SP protein was found in the transcriptomic data. Eight SP genes of *Am. protococcarum* are two quartets of very closely related paralogs that are the result of recent multiplications and belong to the same two orthogroups as both *Am. occidentale* genes. This situation is consistent with the peculiar evolutionary history of *Am. protococcarum* that underwent genome-wide duplications [9].

Two of three SP orthogroups including the genes of Aphelida contain also their orthologs of other opisthokonts, that are fungi, *R. allomyces*, and Holozoa. Obviously, these orthologs are inherited from a common opisthokont ancestor. They retain a structure close to the ancestral one in all opisthokonts and have not undergone multiple duplications (except for *Am. protococcarum*).

The third group of orthologs, containing only the SP genes of aphelids, can hypothetically originate from a gene obtained by an ancestor of aphelids from bacteria by horizontal gene transfer, but its origin has not yet been elucidated.

Separate clades of SP proteins, which are absent in the common ancestor of opisthokonts, appeared in evolution after the isolation of the holozoans and fungi lineages. It can be seen that the specific SP genes of Metazoa, and especially of fungi, have undergone multiple duplications. There are at least 6 reliable specific fungal clades on the constructed tree. Two of them are clearly divided into subclades, each containing proteins from different fungal taxa. Thus, there are about 6–11 specific fungal SP orthogroups. This fact agrees well with the tendency to the enhanced evolution of metabolic genes shown in fungi [20,44].

It is also clear that aphelids, as well as *R. allomyces*, have no fungi-specific SP proteins, similarly, neither aphelids, nor *R. allomyces* have the fungi-specific DHA-2 family of MFS proteins.

The number of SP proteins in aphelids, *R. allomyces* and *C. owczarzewski* is minimal for opisthokonts. For *B. helicus* the lowest number of these proteins compared to other chytrids may be a result of incomplete genome assembly (see Table 1). In addition, the number and diversity of SP proteins in aphelids tend to decrease over the course of aphelid evolution.

All these facts indicate that aphelids do not show at the genomic level any signs of likeness to fungi in enhancing their osmotrophic abilities. Moreover, the osmotrophic abilities of aphelids look even reduced compared to free-living unicellular opisthokonts. One of the reasons could probably be a specialization to the endobiotic lifestyle. While it is advantageous for a free-living cell to have a wide set of trophic possibilities to cope with environmental challenges, an endobiont can get a more specialized feeding mode, since its environment is probably more stable. In the case of aphelids, feeding by phagocytosis turns out to be such a single option. Probably, the same reason may explain a limited repertoire of genes involved in the osmotrophy machinery observed in *R. allomyces* and *C. owczarzewski*.

These results are fully consistent with previously obtained data showing differences in the sets of receptor-like protein kinases and carbohydrate processing enzymes in aphelids and fungi [9]. All these differences suggest that the common ancestor of fungi and aphelids did not have any, even hidden, traces of fungal osmotrophy, which appeared in fungi after the separation of these two lineages.

Hence, ideas about the morphology and lifestyle of the common ancestor of aphelids and fungi become vaguer than just the assumption that it was aphelid-like. Recently, a number of common genes have been discovered in fungi and aphelids [6,8,9,20] and some metabolic features of their common ancestor have been identified [8]. However, if we consider separately each of the common features of fungi and aphelids, they do not allow us to reconstruct precise image of their common ancestor.

1) The noted evidence that the common ancestor of fungi and aphelids fed on algae [8] does not indicate how such feeding occurred. If the cell walls of algae were the original substrate for fungi [45], then the path of transition from endobiotic cytoplasmophagy to extracellular digestion of algal cell walls remains unclear. Modern aphelids, although they have cellulases, do not use extracellular cleavage products and do not show the prerequisites for the formation of a fungal type of nutrition.

In this regard, it is interesting that the possibility of extracellular degradation of polysaccharides was noted in recently discovered organisms from basal lineages of Holozoa [46]. Being cytoplasmophagous predators, these organisms are also capable of bacteriophagy, feeding on particles of solid carbohydrates, and extracellular degradation of carbohydrates. Possibly, the ancestral forms of opisthokonts could also have had a whole range of trophic possibilities, which were further developed in various lineages of this supergroup. Hence, it may be that the common ancestor of aphelids with fungi could be a free-living organism with a wide range of trophic possibilities, rather than a specialized endobiont.

2) In this connection, it is difficult to understand whether the reduction of the actin-associated protein complex CCC in aphelids [9] is a feature originating from a common ancestor with fungi. If the disappearance/reduction of the CCC complex in fungi is associated with the loss of phagocytosis, the reasons for its partial decrease in aphelids are still unknown. This is probably due to the parasitoid feeding of aphelids. If the specific feeding habits of fungi and aphelids were formed after the separation of the lineages, the reduction of the CCC complex could occur independently, in each case for its own specific reason, although the result of the reduction seems to be similar. In fungal ancestry, this course of events is very common [47].

3) The presence of chitin processing enzymes in aphelids and their homology with fungal ones [9], as well as the homology of the protein acting in the infection tube of aphelidian cyst with the hyphal polarization protein [6] reveals an element of cyst germination machinery of the ancestor of aphelids and fungi. But the noted cyst feature does not indicate that it was an infectious agent and, in general, does not say how the cyst functioned in a common ancestor. The cyst with chitinous wall and chitin processing enzymes are characteristic of most unicellular opisthokonts and, obviously, represent symplesiomorphies of opisthokonts [46,48,49].

4) Even the zoosporic life cycle, which is shared by fungi and aphelids, also seems to be a symplesiomorphy of Opisthokonta [48,49]. Moreover, some evidence allows us to raise the question: might it be formed several times in different lineages on the basis of an ancestral ability for cellular polymorphism? This question is possible due to proposal that the genetic basis of the life cycles of fungi and aphelids may be different [50], as well as the recent discovery of basal Holozoa with a high capacity for cellular polymorphism [46], which could be inherited from the common ancestor of Opisthokonta. Such an assumption looks unexpected but will not be so surprising given the wide convergence in morphogenetic processes in fungi [47,51].

Summing up, it turns out that all similar features of fungi and aphelids are either symplesiomorphies, characteristic of all Opisthokonta supergroup, or apomorphies that can be realized in any way of life, or probable evolutionary parallelisms. The wide distribution of convergences and parallelisms in fungi, which complicates and “confuses” the analysis of their evolution, has been repeatedly noted by various authors [47,51]. About the common ancestor of

aphelids with fungi, one can only say that it somehow fed on algae and had either a well-established zoospore life cycle or cellular polymorphism including amoeba, flagellate, and a cyst. Within algae nutrition and polymorphic abilities, there is still insufficient data for further refinement.

It can be reasonably assumed that the aforementioned genes for the chitin processing enzyme and growth tube proteins served as a pre-adaptation to the appearance of the fungal trait complex. However, these cyst-associated genes must have switched to functioning at the vegetative stage, or, conversely, the cyst must have acquired vegetative abilities [8]. In any case, the appearance of morphophysiological features of fungi was obviously based on the evolution of regulatory genes that changed the timing of expression of some components in the chains of gene interaction. At the same time, the prerequisites that made this switch adaptive should have been osmotrophy and extracellular digestion [45], from which fungal evolution should have begun.

In the field of taxonomy, the absence of unambiguous fungal characters in Aphelida makes desirable further discussion on the composition of the taxon Fungi. What are the apomorphies of this taxon after all? Should the diagnosis of Fungi include morphological and physiological characters, or can this taxon be characterized only by a common set of genes and proteins regardless of their functions? Probably, the solution to the question of the inclusion of aphelids and other “early divergent fungal lineages” in the Fungi should be sought in two directions: 1) by studying the set of similarities and differences in fungi and closely related organisms and 2) by discussing the rules of description of the kingdom Fungi. Obviously, for these purposes, further studies on the genomes and the genetic basis of the morphogenesis of aphelids and fungi are required.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: The phylogenetic tree of SP proteins of the studied opisthokont species. Full version; Figure S2: The phylogenetic tree of SP proteins of the studied fungal, aphelid and bacterial species. Full version.

Funding: This work was made in the frame of lab topic 1021051402849-1 ZIN RAS and supported by RSF grant 21-74-20089 (genome sequencing and analysis, manuscript writing and discussion). Cultivation of the aphelids was supported by grant of the Ministry of Science and Higher Education of the Russian Federation 075-15-2021-1069.

Acknowledgments: We thank D.V. Tikhonenkov and V.V. Zlatogursky for discussing the materials and providing various information on the research topic, and the staff of the Zoological Institute of the Russian Academy of Sciences for creating comfortable working conditions. We thank the Centers for Molecular and Cell Technologies and Biobank of Research Park SPbSU, Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency for providing sequencing and computational resources for this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gromov, B.V. Algal parasites of the genera *Aphelidium*, *Amoeboaphelidium* and *Pseudoaphelidium* from the Cienkovski's “Monadea” group as representatives of new class. *Zool. Z.* **2000**, *79*, 517–525.
2. Karpov, S.A.; Mamkaeva, M.A.; Aleoshin, V.V.; Nasonova, E.; Lilje, O.; Gleason, F.H. Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Front. Microbiol.* **2014**, *5*; DOI: 10.3389/fmicb.2014.00112.
3. Letcher, P.M.; Lopez, S.; Schmieder, R.; Lee, P.A.; Behnke, C.; Powell, M.J.; McBride, R.C. Characterization of *Amoeboaphelidium* *protococcarum*, an algal parasite new to the cryptomycota isolated from an outdoor algal pond used for the production of biofuel. *PLoS One* **2013**, *8*(2); DOI: 10.1371/journal.pone.0056232.
4. Adl, S.M.; Bass, D.; Lane, C.E.; Lukes, J.; Schoch, C.L.; Smirnov, A. et al. Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.* **2019**, *66*, 4–119; DOI: 10.1111/jeu.12691.
5. Letcher, P.M.; Powell, M.J. A taxonomic summary of Aphelidiaceae. *IMA Fungus* **2019**, *10*(4); DOI: 10.1186/s43008-019-0005-7.
6. Torruella, G.; Grau-Bové, X.; Moreira, D.; Karpov, S.A.; Burns, J.A.; Sebé-Pedrós, A.; Völcker, E.; López-García, P. Global transcriptome analysis of the aphelid *Paraphelidium tribonemae* supports the phagotrophic origin of fungi. *Commun. Biol.* **2018**, *1*(231); DOI: 10.1038/s42003-018-0235-z.
7. Tedersoo, L.; Sánchez-Ramírez, S.; Kõljalg, U.; Bahram, M.; Döring, M.; Schigel, D.; May, T.; Ryberg, M.; Abarenkov, K. High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Divers.* **2018**, *90*, 135–159; DOI: 10.1007/s13225-018-0401-0.

8. Galindo, L.J.; Torruella, G.; López-García, P.; Ciobanu, M.; Gutiérrez-Preciado, A.; Karpov, S.A.; Moreira, D. Phylogenomics supports the monophyly of aphelids and fungi and identifies new molecular synapomorphies. *Syst. Biol.* **2022**, *syac054*; DOI: 10.1093/sysbio/syac054.
9. Mikhailov, K.V.; Karpov, S.A.; Letcher, P.M.; Lee, Ph.A.; Logacheva, M.D.; Penin, A.A.; Nesterenko, M.A.; Pozdnyakov, I.R.; Potapenko, E.V.; Sherbakov, D.Y.; Panchin, Y.V.; Aleoshin, V.V. Genomic analysis reveals cryptic diversity in aphelids and sheds light on the emergence of Fungi. *Curr. Biol.* **2022**, *32*, 1–13; DOI: 10.1016/j.cub.2022.08.071.
10. Guarro, J.; Gené, J.; Stchigel, A.M. Developments in fungal taxonomy. *Clin Microbiol Rev.* **1999**, *12*(3), 454–500. DOI: 10.1128/CMR.12.3.454.
11. Margulis, L.; Chapman, M.J. Chapter Four - Kingdom Fungi. In *Kingdoms and Domains*, 4th ed; Margulis, L.; Chapman, M.J.; Eds.; Academic Press, 2009, pp. 379–409; DOI: 10.1016/B978-0-12-373621-5.00004-0.
12. Naranjo-Ortiz, M.A.; Gabaldón, T. Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. *Biol. Rev.* **2019**, *94*, 2101–2137; DOI: 10.1111/brv.12550.
13. Richards, T.A.; Leonard, G.; Wideman, J.G. What Defines the “Kingdom” Fungi? *Microbiol. Spectr.* **2017**, *5*(3). DOI: 10.1128/microbiolspec.FUNK-0044-2017.
14. Henderson, P.J. The homologous glucose transport proteins of prokaryotes and eukaryotes. *Res. Microbiol.* **1990**, *141*(3), 316–28; DOI: 10.1016/0923-2508(90)90005-b.
15. Pao, S.S.; Paulsen, I.T.; Saier, M.H., Jr. Major Facilitator Superfamily. *Microbiol. Mol. Biol. R.* **1998**, *62*(1), 1092–2172; DOI: 10.1128/MMBR.62.1.1-34.1998.
16. Yan, N. Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem. Sci.* **2013**, *38*(3), 151–159; DOI: 10.1016/j.tibs.2013.01.003; PMID 23403214.
17. InterPro. Classification of protein families. Major facilitator superfamily domain. Available online: <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR020846/> (accessed on 18.07.2023).
18. Reddy, V.S.; Shlykov, M.A.; Castillo, R.; Sun, E.I.; Saier, M.H. The major facilitator superfamily (MFS) revisited. *FEBS J.* **2012**, *279*(11), 2022–2035; DOI: 10.1111/j.1742-4658.2012.08588.x.
19. Gonçalves, C.; Coelho, M.A.; Salema-Oom, M.; Gonçalves, P. Stepwise Functional Evolution in a Fungal Sugar Transporter Family. *Mol. Biol. Evol.* **2016**, *33*(2), 352–366; DOI: 10.1093/molbev/msv220.
20. Merényi, Z.; Krizsán, K.; Sahu, N.; Liu, X.; Bálint, B.; Stajich, J.; Spatafora, J.W.; Nagy, L.G. Taxonomic vs genomic fungi: contrasting evolutionary loss of protistan genomic heritage and emergence of fungal novelties. *bioRxiv* **2022**, *11.15.516418*; DOI: <https://doi.org/10.1101/2022.11.15.516418>.
21. Van Dijck, P.; Brown, N.A.; Goldman, G.H.; Rutherford, J.; Xue, C.; Van Zeebroeck, G. Nutrient sensing at the plasma membrane of fungal cells. *Microbiol. Spectrum* **2017**, *5*(2); DOI: 10.1128/microbiolspec.FUNK-0031-2016.
22. Bisson, L.F.; Coons, D.M.; Kruckeberg, A.L.; Lewis, D.A. Yeast sugar transporters. *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*(4), 259–308; DOI: 10.3109/10409239309078437.
23. Goffeau, A.; Park, J.; Paulsen, I.T.; Jonniaux, J.L.; Dinh, T.; Mordant, P.; Saier, M.H. Multidrug-resistant transport proteins in yeast: complete inventory and phylogenetic characterization of yeast open reading frames with the major facilitator superfamily. *Yeast* **1997**, *13*(1):43–54; DOI: 10.1002/(SICI)1097-0061(199701)13:1<43::AID-YEA56>3.0.CO;2-J.
24. Dias, P.J.; Sá-Correia, I. The drug:H⁺ antiporters of family 2 (DHA2), siderophore transporters (ARN) and glutathione:H⁺ antiporters (GEX) have a common evolutionary origin in hemiascomycete yeasts. *BMC Genomics* **2013**, *14*, 901; DOI: 10.1186/1471-2164-14-901.
25. Karpov, S.A.; Vishnyakov, A.E.; López-García, P.; Zorina, N.A.; Ciobanu, M.; Tcvetkova, V.S.; Moreira, D. Morphology and molecular phylogeny of *Aphelidium insulamus* sp. nov. (Aphelida, Opisthosporidia). *Protistology* **2020**, *14*(4), 191–203; DOI: 10.21685/1680-0826-2020-14-4-3.
26. Fahrni, J.F.; Bolivar, I.; Berney, C.; Nassonova, E.; Smirnov, A.; Pawlowski, J. Phylogeny of lobose amoebae based on actin and small-subunit ribosomal RNA genes. *Mol. Biol. Evol.* **2003**, *20*(11), 1881–1886; DOI: 10.1093/molbev/msg201.
27. Kolmogorov, M.; Yuan, J.; Lin, Yu; Pevzner, P. Assembly of Long Error-Prone Reads Using Repeat Graphs. *Nat. Biotechnol.* **2019**, *37*(5), 540–546; DOI: 10.1038/s41587-019-0072-8.
28. Challis, R.; Richards, E.; Rajan, J.; Cochrane, G.; & Blaxter, M. BlobToolKit—interactive quality assessment of genome assemblies. *G3-Genes Genom. Genet.* **2020**, *10*(4), 1361–1374; DOI: 10.1534/g3.119.400908.
29. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinform.* **2018**, *34*(18), 3094–3100; DOI: 10.1093/bioinformatics/bty191.
30. Vaser, R.; Sović, I.; Nagarajan, N.; Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **2017**, *27*(5), 737–746; DOI: 10.1101/gr.214270.116.
31. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; Earl, A.M. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **2014**, *19*;9(11):e112963; DOI: 10.1371/journal.pone.0112963.

32. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinform.* **2015**, *31*(19), 3210–3212; DOI: 10.1093/bioinformatics/btv351.
33. Gurevich, A.; Saveliev, V.; Vyahhi, N.; & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinform.* **2013**, *29*(8), 1072–1075; DOI: 10.1093/bioinformatics/btt086.
34. Palmer, J.; Stajich, J. Nextgenusfs/funannotate: funannotate v1.8.1 (Version 1.8.1). *Zenodo* **2020**; DOI: 10.5281/zenodo.4054262.
35. Haas, B.J.; Salzberg, S.L.; Zhu, W.; Pertea, M.; Allen, J.E.; Orvis, J.; White, O.; Buell, C.R.; Wortman, J.R. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **2008**, *9*, R7; DOI: 10.1186/gb-2008-9-1-r7.
36. Käll, L.; Krogh, A.; Sonnhammer, E.L. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res.* **2007**, *35* (Web Server issue), W429–432; DOI: 10.1093/nar/gkm256.
37. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. In: *Gene Prediction. Methods in Molecular Biology*; Kollmar, M; Ed.; 2019; Humana: New York, NY; Volume 1962; DOI: 10.1007/978-1-4939-9173-0_1.
38. Jones, P.; Binns, D.; Chang H-Yu, Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; Pesseat, S.; Quinn, A.F.; Sangrador-Vegas, A.; Scheremetjew, M.; Yong S-Y, Lopez, R.; Hunter, S. InterProScan 5: genome-scale protein function classification. *Bioinform.* **2014**, *30*(9), 1236–1240; DOI: 10.1093/bioinformatics/btu031.
39. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer K & Madden, T.L. BLAST+: architecture and applications. *BMC Bioinform.* **2009**, *10*, 421; DOI: 10.1186/1471-2105-10-421.
40. Wheeler, T.J.; Eddy, S.R. nhmmer: DNA homology search with profile HMMs. *Bioinform.* **2013**, *29*(19), 2487–2489; DOI: 10.1093/bioinformatics/btt403.
41. Di Tommaso, P.; Moretti, S.; Xenarios, I.; Orobitz, M.; Montanyola, A.; Chang, J.M.; Taly, J.F.; Notredame, C. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **2011**, *39* (Web Server issue), W13-7; DOI: 10.1093/nar/gkr245.
42. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinform.* **2009**, *25*, 1972–1973; DOI: 10.1093/bioinformatics/btp348.
43. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*(1), 268–274; DOI: 10.1093/molbev/msu300.
44. Ocaña-Pallarès, E.; Williams, T.A.; López-Escardó, D.; Arroyo, A.S.; Pathmanathan, J.S.; Baptiste, E.; Tikhonenkov, D.V.; Keeling, P.J.; Szöllősi, G.J.; Ruiz-Trillo, I. Divergent genomic trajectories predate the origin of animals and fungi. *Nature* **2022**, *609*, 747–753; DOI: 10.1038/s41586-022-05110-4.
45. Chang, Y.; Wang, S.; Sekimoto, S.; Aerts, A.L.; Choi, C.; Clum, A.; LaButti, K.M.; Lindquist, E.A.; Ngan CYe, Ohm, R.A.; Salamov, A.A.; Grigoriev, I.V.; Spatafora, J.W.; Berbee, M.L. Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. *Genome Biol. Evol.* **2015**, *7*(6), 1590–1601; DOI: 10.1093/gbe/evv090.
46. Tikhonenkov, D.V.; Hehenberger, E.; Esaulov, A.S.; Belyakova, O.I.; Mazei, Y.A.; Mylnikov, A.P.; Keeling, P.J. Insights into the origin of metazoan multicellularity from predatory unicellular relatives of animals. *BMC Biol.* **2020**, *18*(39); DOI: 10.1186/s12915-020-0762-1.
47. Brun, S; Silar, P. Convergent evolution of morphogenetic processes in Fungi. In: *Evolutionary biology – Concepts, molecular and morphological evolution*; Pontarotti, P.; Ed.; Springer: Berlin, Heidelberg, 2010; pp. 317–328; DOI: 10.1007/978-3-642-12340-5_19.
48. Kożyczkowska, A.; Najle, S.R.; Ocaña-Pallarès, E.; Aresté, C.; Shabardina, V.; Ara, P.S.; Ruiz-Trillo, I.; Casacuberta, E. Stable transfection in protist *Corallochytrium limacisporum* identifies novel cellular features among unicellular animals relatives. *Curr. Biol.* **2021**, *31*(18), 4104–4110.e5; DOI: 10.1016/j.cub.2021.06.061.
49. Torruella, G.; de Mendoza, A.; Grau-Bové, X.; Antó, M.; Chaplin, M.A.; del Campo, J.; Eme, L.; Pérez-Cordón, G.; Whipps, C.M.; Nichols, K.M.; Paley, R.; Roger, A.J.; Sitjà-Bobadilla, A.; Donachie, S.; Ruiz-Trillo, I. Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Curr. Biol.* **2015**, *25*(18), 2404–2410; DOI: 10.1016/j.cub.2015.07.053.
50. Pozdnyakov, I.R.; Zolotarev, A.V.; Karpov, S.A. Comparative analysis of zoosporogenesis' genes of the blastoclad *Blastocladiella emersonii* and the aphelid *Paraphelidium tribonematis* reveals the new directions of evolutionary research. *Protistology* **2021**, *15*(1), 10–23; DOI: 10.21685/1680-0826-2021-15-1-2.
51. Nagy, L.G.; Kovács GM and Krizsán, K. Complex multicellularity in fungi: evolutionary convergence, single origin, or both? *Biol. Rev. Camb. Philos. Soc.* **2018**, *93*(4), 1778–1794; DOI: 10.1111/brv.12418.