DR. NATALIIA ANNENKOVA (Orcid ID : 0000-0001-6969-9938)

Article type : Note

Delineating closely related dinoflagellate lineages using phylotranscriptomics [1]

Nataliia V. Annenkova[2]

Limnological Institute Siberian Branch of the Russian Academy of Sciences 3, Ulan-Batorskaya St., 664033, Irkutsk, Russia

Dag Ahrén

Microbial Ecology Group, Department of Biology, Lund University, Ecology Building, SE-223 62 Lund, Sweden

Bioinformatics Infrastructures for Life Sciences (BILS), Department of Biology, Lund University, Ecology Building, SE-223 62 Lund, Sweden

Ramiro Logares

Department of Marine Biology and Oceanography, Institute of Marine Science (ICM)–Consejo Superior de Investigaciones Científicas (CSIC), Passeig Marítim de la Barceloneta 37-49, Barcelona E08003, Spain

Anke Kremp

Marine Research Centre, Finnish Environment Institute, Erik Palmenin aukio 1, 00560 Helsinki, Finland

Karin Rengefors

Aquatic Ecology, Department of Biology, Lund University, Ecology Building, SE-223 62 Lund, Sweden

Corresponding author: tasha.annenkova@gmail.com, +79025138236

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jpy.12748

Running title: Phylotranscriptomics of closely related dinoflagellates

Editorial Responsibility: P. Gabrielson (Associate Editor)

Abstract

Recently radiated dinoflagellates *Apocalathium aciculiferum* (collected in Lake Erken, Sweden), *A. malmogiense* (Baltic Sea) and *Apocalathium* aff. *malmogiense* (Highway Lake, Antarctica) represent a lineage with an unresolved phylogeny. We determined their phylogenetic relationships using phylotranscriptomics based on 792 amino acid sequences. Our results showed that *A. aciculiferum* diverged from the other two closely related lineages, consistent with their different morphologies in cell size, relative cell length and presence of spines. We hypothesized that *A.* aff. *malmogiense* and *A. malmogiense*, which inhabit different hemispheres, are evolutionarily more closely related because they diverged from a marine common ancestor, adapting to a wide salinity range, while *A. aciculiferum* colonized a freshwater habitat, by acquiring adaptations to this environment, in particular, salinity intolerance. We show that phylotranscriptomics can resolve the phylogeny of recently diverged protists. This has broad relevance, given that many phytoplankton species are morphologically very similar, and single genes sometimes lack the information to determine species' relationships.

Key index words: adaptive radiation; High Throughput Sequencing; microalgae; protists; phylogenomics, transcriptome

Abbreviations: ITS2, internal transcribed spacer-2; COB, cytochrome b; BS, bootstrap support

To date, phylogenomic studies of microbial eukaryotes have focused mainly on resolving deep evolutionary relationships among major taxa (e.g., Burki et al. 2012, Janouškovec et al. 2017), while few studies have focused on recently diverged species/lineages (e.g., Gayevskiy and Goddard 2016). However, phylogenomics can potentially help in recovering evolutionary relationships among close relatives: compared to single-gene approaches, phylogenomics may have more resolution, given that the phylogenetic information of many genes is used for tree reconstruction (McCormack et al. 2013). Transcriptomic data is a promising resource for such analyses, since they contain a set of expressed genes, which can be used as orthologs. Transcriptomes are also less expensive to sequence using High Throughput Sequencing, as well as easier to assemble than whole genomes (e.g., Keeling et al. 2014). However, problems in determining evolutionary relationships may occur even with using phylogenomics, because of various non-phylogenetic signals (see Philippe et al. 2011, Som 2014).

Here we use phylotranscriptomic analyses to determine the evolutionary relationships among closely related dinoflagellate lineages from the genus *Apocalathium*. These lineages constitute a recently radiated species complex with unresolved phylogeny that displays both cryptic diversity (that is, morphological similarity despite genetic differences) and, at the same time, diverse phenotypes in different environments (Annenkova et al. 2015). A particular feature of this complex is that despite having four different morphospecies, these taxa have very small differences in the 772 bp internal transcribed spacer-2(ITS2)-LSU rRNA gene fragments (up to 0.9%) and in a 592 bp cytochrome b (COB) gene fragments (0 – 1.7%; Annenkova et al. 2015). The common range in genetic variation between dinoflagellate species is 3.6–41.5% for ITS (Stern et al. 2012) and more than 0.89% for a 385 bp COB fragment (Lin et al. 2008) that overlaps the 592 bp fragment . The specific goal of this study was to resolve the relationships among three (those available as clonal cultures) of the key

members in the *Apocalathium* complex: *A. aciculiferum* (≡ *Peridinium aciculiferum*, represented by strain PAER2 from the lake Erken, Sweden), *A. malmogiense (≡ Scrippsiella hangoei*, SHTV5 from the Northern Baltic Sea) and *A.* aff. *malmogiense (≡ Scrippsiella* aff. *hangoei*, SHHI from the brackish Highway Lake, Antarctica). *Apocalathium malmogiense* and *A.* aff. *malmogiense* have virtually identical morphologies (some variations were noted in the shape of plates), but they have law differences in the rRNA genes and ITS2 spacer (Logares et al. 2008). Both lineages also occur in similar conditions (cold, brackish water), but in geographically very distant locations, the Baltic Sea and an Antarctic lake (Rengefors et al. 2008). As the Baltic lineage was initially described as *A. malmogiense,* the term affinis (aff.) was used for the Antarctic lineage, which has both morphological and genetic affinity to *A. malmogiense*, but it is not identical (Rengefors et al. 2008). *Apocalathium aciculiferum* differs from *A. malmogiense* in morphology (*A. aciculiferum* is longer, wider, more elongated and normally displays three to four spines that are absent in *A. malmogiense*) and salinity tolerance, but they are identical in their rRNA genes and ITS2 spacer sequences (Logares et al. 2007). COB genes differ in all three lineages and even in different populations of the lineages (Logares et al. 2008).

Representative strains were grown as clonal non-axenic cultures isolated from single-cells. Total RNA was extracted and sequenced using the Illumina HiSeq2000 platform. De novo transcriptomes from each lineage were assembled and included in phylogenomic analyses together with corresponding data from other dinoflagellates that were downloaded from the iMicrobe Project (http://data.imicrobe.us) and NCBI. Even though using amino acid sequences during phylogenetic reconstruction instead of DNA sequences may reduce variability, we used this approach as amino acid translations decrease the influence of non-ortholog gene copies on the phylogenetic signal (Bachvaroff et al. 2014). Moreover, large datasets such as ours should provide sufficient numbers of substitutions at the protein level.

The effects of possible paralogs in the alignment were evaluated, the final Maximum-Likelihood tree was built and the stability of its topology was assessed (detailed methods in Appendix S1 in the Supporting Information). *Prorocentrum cordatum* was used as the outgroup (Orr et al. 2012). The sequences obtained in this study have been deposited in the iMicrobe Project (MMETSP0359-MMETSP61, MMETSP0367 - MMETSP0371), and the final tree and its alignment were deposited in TreeBase (http://purl.org/phylo/treebase/phylows/study/TB2:S21510).

The final data set contained 792 groups of protein orthologs that corresponded to 212,219 amino acid characters (all characters present for the 11 species). There were 114,894 sites with variation in at least one taxon; of them 70,910 sites were parsimony informative. Our lineages of interest (*Apocalathium aciculiferum, A. malmogiense and A.* aff. *malmogiense*) were very closely related, and thus they displayed low levels of substitutions among themselves: there were 5,680 variable sites in their sequences and within them about 1,200 sites were informative. The distribution of these sites was not equal among single sequences: 51 sequences contained more than 19 variable sites, 221 sequences had from 5 to 19 variable sites, 432 sequences had from one to five variable sites and 88 sequences were identical for all three lineages (Fig. S1 in the Supporting Information).

The phylogenetic analysis based on all 792 amino acid fragments resulted in a tree, where all clades had high bootstrap support (BS; Fig. 1B). Its topology was largely identical to the tree that was obtained based on partial SSU and LSU rRNA genes (Fig. 1A), but differed within *Apocalathium*. In the phylotranscriptomics tree (Fig. 1B), *A. malmogiense* formed a clade with Antarctic *A.* aff. *malmogiense*, and *A. aciculiferum* was a sister lineage to this clade with 100% BS.

This also differed from the result obtained by Logares et al. (2008) using 818 bp COB gene fragments, where *A. aciculiferum* and *A. malmogiense* clustered together with weak statistical

support.

Despite the high bootstrap support obtained in our phylotranscriptomics tree (Fig. 1B), such a tree based on concatenated alignments may not necessarily reflect the true species tree (see Kubatko and Degnan 2007, Som 2014). Therefore, we explored consistency/discrepancy in our phylogeny to increase the confidence in our species tree estimation (as recommended in Blom et al, 2016). Our analysis of the multiprotein trees based on alignments with different lengths (Fig. 2, Fig. S2 in the Supporting Information) showed that the use of > 200 amino acid sequences from the 792 ortholog groups resulted in the same tree topology as the phylotranscriptomics tree (Fig. 1B). Moreover, the majority rule consensus-tree based on all single-sequence trees had the same topology as the concatenated phylotranscriptomics tree. However, "internode certainty" support, which is used to evaluate trees' incongruences (Salichos et al. 2014), was relatively low (= 0.21) for the *Apocalathium malmogiense* + *A.* aff. *malmogiense* clade, because 36% of single-sequence trees showed alternative topologies. Such a level of incongruence is not considered to be high, and its existence was expected. When speciation events happen in short time spans (like thousands years for an evolutionary radiation), various population processes can have a large influence on the evolutionary relationships (Rosenberg 2002, Knowles and Chan 2008). For example, polymorphisms could be randomly fixed in recently diverged species, and/or they could feature different ancestral polymorphisms due to incomplete lineage sorting (e.g., Pollard et al. 2006). We suggest that some of our trees showing an alternative topology is due to such population level processes. This applies especially to the incongruent trees with high statistical support for the alternative topology. However, many of the single-sequence trees, which contain alternative topologies, could be the result of methodological/stochastic artifacts. On the one hand, proteins in the *Apocalathium* complex showing few or no substitutions (Fig. S1D) may yield alternative topologies due to lack of phylogenetic signal. On the other hand, the fastest-

evolving sequences may yield alternative topologies due to mutational saturation and long-branch attraction artifacts (e.g., Léveillé-Bourret et al. 2017). All but two trees with more than 50 substitutions in sequences within our lineages (Fig. S1A) exhibited topologies in which the three studied lineages were not monophyletic. Overall, it is clear, that the phylotranscriptomics tree based on 792 amino acid sequences represents the main evolutionary trend in the studied complex, and we regard it as the species tree.

According to the phylotranscriptomics tree (Fig. 1B) *Apocalathium* aff. *malmogiense* from the Antarctic brackish lake and *A. malmogiense* from the Baltic Sea are evolutionary closer to each other than to *A. aciculiferum*, isolated from the lake Erken near the Baltic Sea coast. This conclusion was corroborated by morphological data. *Apocalathium malmogiense* and *A.* aff. *malmogiense* have almost the same morphology and cannot be differentiated microscopically (Fig. 1, C and D), whereas *A. aciculiferum* (Fig. 1E) differs in general morphology from both. It is bigger, more elongated and displays several spines, but the plate pattern of *A. aciculiferum* and *A. malmogiense* is identical (Logares et al. 2007, Annenkova et al. 2015). Recent analysis of their plastid genes also showed, that *A. malmogiense* and *A.* aff. *malmogiense* have a common ancestor with specific in-frame deletion and loss of otherwise conserved residues from the *atp*A gene (Dorrel et al. 2016). We hypothesize that *A.* aff. *malmogiense* and *A. malmogiense*, which inhabit different hemispheres, cluster together mainly because they have diverged from a marine common ancestor adapted to a wide salinity range, ranging from 0 to 30 (Logares et al. 2007a, this study). *Apocalathium aciculiferum* colonized the freshwater habitat and lost the ability to grow at salinities above 3 (Logares et al. 2007a). Indeed, *A. aciculiferum* differs from the other two members of the genus mainly in the proteins featuring functions related to environmental responses (e.g., processes associated with protein modifications, transport and targeting to membranes, regulation of the cell signal transduction, ceramide metabolism, RNA processing; Fig. S3 in

the Supporting Information). Additionally, despite the large geographical distance, gene flow between Baltic and Antarctic lineages cannot be totally excluded. Evidence of Pole-to-Pole gene flow has been established for ciliates (Di Giuseppe et al. 2013), foraminiferans (Darling et al. 2000) and acanthoecid choanoflagellates (Nitsche and Arndt 2015). Yet, different salinity regimes are expected to limit interbreeding between *A. aciculiferum* and *A. malmogiense*, although these lineages are located geographically close to each other (Logares et al. 2007b, 2009).

Another possible explanation for the observed phylogenetic relationships is convergent evolution due to similar salinity conditions, i.e., similar phenotypes that evolved in parallel (Simpson 1961). However, only extreme genetic parallelism (the same mutations in the same places) across lineages can potentially result in the clustering of *A. malmogiense* and *A.* aff. *malmogiense* together, which is highly unlikely. The similarity between two closely related species is typically the product of common ancestry, rather than parallel evolution (Elmer and Meyer 2011). Hence, we consider that parallel evolution is a highly improbable explanation for the phenotypic and genotypic similarity of *A. malmogiense* and *A.* aff. *malmogiense*.

Overall, we conclude, that our phylotranscriptomics tree has resolved the relationships among the recently radiated lineages. Our phylogeny also provides some indication that *A. aciculiferum* diverged from the other lineage due to adapting to a different environmental condition, different salinity. Future studies including more lineages from *Apocalathium* (in particular, *A. baicalense* and *A. euryceps*) are needed to prove this. Today most comprehensive studies on the recent evolutionary radiation have so far focused on animals (e.g., Wagner et al. 2012) or plants (e.g., Hughes and Eastwood 2006), and there are many open questions regarding this topic in microbial eukaryotes (e.g., what environmental factors play more important roles in such processes; why do some species radiate and others do not; how often does radiation takes place within protist lineages). Further studies using a

genomics approach should help resolve these questions.

Acknowledgments

The work was supported through the Russian Foundation for Basic Research (project # 16-04-01704) to NA and a Swedish Research Council grant (2012-10-24) to KR as well as the marine microbial eukaryote transcriptome sequencing project (MMETSP) of the Gordon and Betty Moore Foundation. Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) (project No. b2013085). We thank Dr. Purificación López-García, Dr. Michael Cummings, the Editor and two anonymous reviewers for invaluable advice. Marie Svensson and Karolina Härnström are thanked for laboratory assistance. The authors declare no conflict of interest.

References

Annenkova, N. V., Hansen, G., Moestrup, Ø. & Rengefors, K. 2015. Recent adaptive radiation in a marine and freshwater dinoflagellate species flock. *ISME J.* 9:1821-34.

Bachvaroff, T. R., Gornik, S. G., Concepcion, G. T., Waller, R. F., Mendez, G. S., Lippmeier, J. C. & Delwiche, C. F. 2014. Dinoflagellate phylogeny revisited: using ribosomal proteins to resolve deep branching dinoflagellate clades. *Mol. Phylogenet. Evol.* 70:314-22.

Blom, M. P. K., Bragg, J. G., Potter, S. & Moritz, C. 2017. Accounting for uncertainty in gene tree estimation: summary-coalescent species tree inference in a challenging radiation of Australian lizards. *Syst. Biol.* 66:352-66.

Burki, F., Shalchian-Tabrizi, K., Minge, M., Skjæveland, Å., Nikolaev, S. I., Jakobsen, K. S. et al. 2007. Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE* 2:e790.

Craveiro, S. C, Daugbjerg, N., Moestrup, Ø. & Calado, A. J. 2017. Studies on *Peridinium aciculiferum* and *Peridinium malmogiense* (=*Scrippsiella hangoei*): comparison with *Chimonodinium lomnickii* and description of *Apocalathium* gen. nov. (Dinophyceae). *Phycologia* 56:21-35.

Darling, K. F., Wade, C. M., Stewart, I. A., Kroon, D., Dingle, R. & Brown, A. J. L. 2000. Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Nature* 405:43–47.

Di Giuseppe, G., Barbieri, M., Vallesi, A., Luporini, P. & Dini, F. 2013. Phylogeographical pattern of *Euplotes nobilii*, a protist ciliate with a bipolar biogeographical distribution. *Mol. Ecol.* 22:4029–37.

Dorrell, R.G., Klinger, C.M., Newby, R.J., Butterfield, E.R., Richardson, E., Dacks, J.B., Howe C. J. et al. 2017. Progressive and biased divergent evolution underpins the origin and diversification of peridinin dinoflagellate plastids. *Mol. Biol. Evol.* 34:361-79.

Elmer, K. R. & Meyer, A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol. Evol*. 26: 298–306.

Gayevskiy, V. & Goddard, M. R. 2016. *Saccharomyces eubayanus* and *Saccharomyces arboricola* reside in North Island native New Zealand forests. *Environ. Microbiol.* 18:1137–47.

Hughes, C. & Eastwood, R. 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proc. Natl. Acad. Sci. USA* 103:10334-39.

Janouškovec, J., Gavelis, G. S., Burki, F., Dinh, D., Bachvaroff, T. R., Gornik, S. G., Bright et al. 2017. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proc. Natl. Acad. Sci. USA* 114:E171-80.

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., Armbrust, E. V. et al. 2014. The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol*. 12:e1001889.

Knowles, L. L. & Kubatko, L. S. 2010. Estimating species trees: an introduction to concepts and models. *In* Knowles, L. L. & Kubatko, L. S. [Eds.] *Estimating Species Trees: Practical and Theoretical Aspects.* Wiley-Blackwell, Hoboken, New Jersey, pp. 1-14.

Kubatko, L. S. & Degnan, J. H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* 56:17-24.

Léveillé-Bourret, É., Starr, J. R., Ford, B. A., Lemmon, E. M. & Lemmon, A. R. 2018. Resolving rapid radiations within angiosperm families using anchored phylogenomics. *Syst. Biol*. 67:94-112.

Lin, S., Zhang, H., Hou, Y., Zhuang, Y., & Miranda, L. 2009. High-level diversity of dinoflagellates in the natural environment, revealed by assessment of mitochondrial cox1 and

cob genes for dinoflagellate DNA barcoding. *Appl. Environ. Microbiol.* 75:1279–90.

Logares, R., Bråte, J., Bertilsson, S., Clasen, J. L., Shalchian-Tabrizi, K. & Rengefors, K. 2009. Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol.* 17:414-22.

Logares, R., Daugbjerg, N., Boltovskoy, A., Kremp, A., Laybourn-Parry, J. & K. Rengefors. 2008. Recent evolutionary diversification of a protist lineage. *Environ. Microbiol.* 10:1231-43.

Logares, R., Rengefors, K., Kremp, A., Shalchian-Tabrizi, K., Boltovskoy, A., Tengs, T., Shurtleff, A. & Klaveness, D. 2007a. Phenotypically different microalgal morphospecies with identical ribosomal DNA: a case of rapid adaptive evolution? *Microb. Ecol.* 53:549-61.

Logares, R., Shalchian-Tabrizi, K., Boltovskoy, A. & Rengefors, K. 2007b. Extensive dinoflagellate phylogenies indicate infrequent marine-freshwater transitions. *Mol. Phyl. Evol.* 45:887–903.

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C. & Brumfield, R. T. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol. Phyl. Evol*. 66: 526–38.

Nitsche, F. & Arndt, H. 2015. Comparison of similar Arctic and Antarctic morphotypes of heterotrophic protists regarding their genotypes and ecotypes. *Protist* 166:42–57.

Orr, R. J. S., Murray S. A., Stüken A., Rhodes L. & Jakobsen K. S. 2012. When naked

became armored: an eight-gene phylogeny reveals monophyletic origin of theca in dinoflagellates. P*LoS ONE* 7:e50004.

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., & D. Baurain. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* 9:e1000602.

Pollard, D. A., Iyer, V. N., Moses, A. M. & Eisen, M. B. 2006. Widespread discordance of gene trees with species tree in *Drosophila:* evidence for incomplete lineage sorting. *PLOS Genet.* 2:e173.

Rengefors, K., Laybourn-Parry, J., Logares, R., Marshall, W. A. & Hansen, G. 2008. Marine-derived dinoflagellates in Antarctic saline lakes: annual dynamics and community composition. *J. Phycol.* 44:592–604.

Rosenberg, N. A. 2002. The probability of topological concordance of gene trees and species trees. *Theor. Populat. Biol.* 61:225–47.

Salichos, L., Stamatakis, A. & Rokas, A. 2014. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. *Mol Biol Evol.* 31:1261-71.

Simpson, G. G. 1961. *Principles of animal taxonomy*. Columbia University Press, New York, 247 pp.

Som, A. 2015. Causes, consequences and solutions of phylogenetic incongruence. *Brief. Bioinform.* 16:536–48.

Stern, R. F., Andersen, R. A., Jameson, I., Küpper, F. C., Coffroth, M. A. et al. 2012. Evaluating the ribosomal Internal Transcribed Spacer (ITS) as a candidate dinoflagellate barcode marker. *PLOS ONE* 7:e42780.

Wagner, C. E., Harmon, L. J. & Seehausen, O. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* 487:366–69.

Figure 1. Maximum-Likelihood trees based on partial SSU rRNA and LSU rRNA genes (A) and on a 792-protein alignment (B). Studied species are in bold. Dotted lines indicate a clade with short branches containing *Apocalathium aciculiferum* and *A. malmogiense*. Numbers at nodes represent bootstrap values (based on 1000 replicates). C – E. SEM micrographs of: C – *Apocalathium malmogiense* (Baltic Sea, picture from Kremp et al. 2005), D – *Apocalathium* aff. *malmogiense* (Antarctic lake), E – *Apocalathium aciculiferum* (Lake Erken).

Figure 2**.** The graph shows the dependency between fraction of trees that are identical to phylotranscriptomics tree (Fig. 1B) and the number of amino acid sequences that were used in the alignment. X axis positions of the black boxes correspond to the number of ortholog amino acid sequences (10, 25, 50, 100, 150, 175, 200, 225, 250, 300 and 350 sequences) that were picked randomly 50 times from the dataset of 792 orthologs, and each was concatenated into one alignment. Maximum-Likelihood trees were built based on such alignments. Y axis positions show percentage of the trees in each group that were identical to phylotranscriptomics tree in Figure 1B.

Appendix_S1. Materials and methods.

Figure S1. Amino-acids from the studied sequences: black – constant, green – variable parsimony-uninformative, blue – parsimony-informative, red – variable within aciculiferum/malmogiense clade.

Figure S2. Midpoint-rooted Maximum-Likelihood trees with topologies different from the phylotranscriptomic tree topology (Fig. 1B in the main text), divergent branch topology is in red. The trees were constructed based on alignments with 10 randomly picked amino acid sequences and 100 rounds of rapid bootstrap analysis. Trees in panel A and B were the most common among such variants. With increasing number of randomly selected proteins, the topologies converge towards the fully resolved 792-amino acid sequences phylogeny (see Fig. 2 in the main text).

Figure S3. Results of REViGO semantic analysis of GO biological process terms. Functional enrichment of GO-terms was analyzed in the annotated transcripts with using InterProScan and summarized using REVIGO. The terms remaining after the redundancy reduction are represented as scatterplots, where the more semantically similar terms are positioned closer together. A – analysis of the proteins resulted in the phylogenetic trees, which are congruent to the phylotranscriptomic tree (Fig 2B). Light blue bubble show the process which is associated with the main one (e.g., bubble named “cellular protein modification process” is bound with four light blue bubbles “peptidyl-amino acid modification”, “protein peptidyl-prolyl isomerization”, “protein deubiquitination“ and “proteolysis”). B - analysis of the proteins, that are identical within Apocalathium complex, and the proteins resulted in the trees which incongruent to the phylotranscriptomic tree (Fig. 2B).