ISSN 0006-2979, Biochemistry (Moscow), 2015, Vol. 80, No. 11, pp. 1492-1499. © Pleiades Publishing, Ltd., 2015.
Original Russian Text © K. V. Mikhailov, D. V. Tikhonenkov, J. Janouškovec, A. Y. Diakin, M. V. Ofitserov, A. P. Mylnikov, V. V. Aleshin, 2015, published in Biokhimiya, 2015, Vol. 80, No. 11, pp. 1715-1723.

Primary Structure of 28S rRNA Gene Confirms Monophyly of Free-Living Heterotrophic and Phototrophic Apicomplexans (Alveolata)

K. V. Mikhailov[1], D. V. Tikhonenkov[2,3], J. Janouškovec[3,4], A. Y. Diakin[5], M. V. Ofitserov[6], A. P. Mylnikov[2], and V. V. Aleshin[1]*

[1]*Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 119991 Moscow, Russia; fax: +7 (495) 939-3181; E-mail: kv.mikhailov@belozersky.msu.ru; Aleshin@genebee.msu.su*
[2]*Papanin Institute for Biology of Inland Waters, Russian Academy of Sciences, 152742 Borok, Russia; fax: +7 (48547) 24-042; E-mail: tikho-denis@yandex.ru; mylnikov@ibiw.yaroslavl.ru*
[3]*University of British Columbia, Department of Botany, V6T 1Z4 British Columbia, Vancouver, Canada; E-mail: janjan.cz@gmail.com*
[4]*Canadian Institute for Advanced Research, ON M5G 1Z8 Toronto, Canada*
[5]*Masaryk University, Faculty of Science, Department of Botany and Zoology, 60177 Brno, Czech Republic; E-mail: ondredyakin@gmail.com*
[6]*State Scientific Institute of Irrigation Fish Breeding, 142460 Vorovskogo Settlement, Noginsk District, Moscow Region, Russia; fax: + 7 (496) 513-7588; E-mail: miha-akademik@ya.ru*

Received July 7, 2015

Abstract—Phylogenetic analysis of large subunit ribosomal RNA (LSU rRNA or 28S rRNA) gene sequences from free-living predatory flagellates *Colpodella angusta*, *Voromonas pontica*, and *Alphamonas edax* (Apicomplexa) confirms their close relationship with chromerids *Chromera velia* and *Vitrella brassicaformis*, which possess a functional photosynthetic plastid. Together these organisms form a sister group to parasitic apicomplexans (coccidians and gregarines, or sporozoans *sensu lato*). This result agrees with the previous conclusion on monophyly of colpodellids and chromerids (chrompodellids) based on phylogenomic data. The revealed relationships demonstrate a complex pattern of acquisition, loss, or modification of plastids and transition to parasitism during alveolate evolution.

DOI: 10.1134/S0006297915110115

Key words: molecular phylogenetics, ribosomal RNA genes, apicomplexans, Sporozoa, heterotrophic flagellates, plastids, evolution of parasitism

The Sporozoa – intracellular parasites of animals that cause such diseases as malaria, toxoplasmosis, cryptosporidiosis, and others – were traditionally a part of zoology courses. However, comparison of the rRNA nucleotide sequences revealed a close relationship of Sporozoa with algae – testaceous flagellates (dinoflagellates) [1], which raised the question of the evolutionary transition between autotrophic and heterotrophic forms in the alveolate group [2]. Moreover, the discovery of a rudimentary form of plastid (apicoplast) in Sporozoa [3] involved, as is known now, in metabolic processes not related to photosynthesis showed that Sporozoa evolved from an autotrophic ancestor. Investigation of biodiversity of protozoan flagellates led to addition of free-living predatory colpodellids [4, 5] to a group uniting parasitic Sporozoa and autotrophic dinoflagellates. Colpodellids feed on other protists by taking up the content of prey cells using an apical complex. The resemblance of the apical complex in parasites and free-living predators suggests that they inherited this part of cellular organization from a common heterotrophic ancestor [5, 6]. Recently, autotrophic relatives of Sporozoa have been discovered that are more closely related to Sporozoa than dinoflagellates. These autotrophic organisms have been named chromerids: *Chromera velia* [7] and *Vitrella brassicaformis* [8]. They contain functional plastids that potentially share common origin with the sporozoan apicoplast. In addition to the characterized species, the DNA fragments originating from unidentified relatives of either colpodellids or chromerids are being found in metagenomes from various environmental samples. Although the relation of

Abbreviations: OTUs, operational taxonomic units.

* To whom correspondence should be addressed.

colpodellids and chromerids with Sporozoa has been demonstrated in a sufficiently conclusive way, the relationship between them was recently a subject of alternative hypotheses [9-11] and needed clarification. Recently obtained transcriptome data provides substantial support for the hypothesis of monophyly of colpodellids and chromerids and for the possibility of giving them the common name of chrompodellids [12]. In this work, we confirm monophyly of the indicated clade and its sister relation to Sporozoa using a traditional phylogenetic marker, the 28S rRNA gene, as well as combined nucleotide sequences of 18S and 28S rRNA genes.

MATERIALS AND METHODS

Cell cultures and DNA sequencing. Predatory colpodellids *Colpodella angusta* (isolate Spi-2) and *Voromonas pontica* (isolate G-3) were cultivated using bacteriotrophic kinetoplastid flagellates *Parabodo caudatus* (isolate BAS-1) and *Procryptobia sorokini* (isolate B-69) as prey, respectively; *Alphamonas edax* (isolate BE-2) was cultivated on heterotrophic golden algae *Spumella* sp. (isolate OF-40). Prey cells were grown on a suspension of bacteria *Pseudomonas fluorescens*. The sources of cultures and details of cultivation were described earlier [11]. DNA was isolated from *V. pontica* and *A. edax* cultures using the Diatom kit (Izogen, Russia). Overlapping fragments of the 28S rRNA gene were amplified using an Encyclo PCR kit (Evrogen, Russia) and a set of primers [13]. PCR products were cloned into a pTZ57R plasmid vector (Fermentas, Lithuania). The sequences corresponding to fragments of prey and predator genes were selected from clones based on fragment length polymorphism with restriction endonuclease TaqI. Nucleotide sequences of ribosomal genes were sequenced on an Applied Biosystems 3730 DNA Analyzer capillary sequencer (Life Technologies, USA) in the Center of Collective Use at the Engelhardt Institute of Molecular Biology, Russian Academy of Sciences. RNA was isolated from *C. angusta* culture, and on its basis, cDNA was synthesized using a SMARTer Pico PCR cDNA Synthesis kit (Clontech, USA). The cDNA library was sequenced on a HiSeq2000 (Illumina Inc., USA), and contigs were assembled in Inchworm (Trinity v. r2012-06-08).

Selection of sequences and phylogenetic analysis. A search for nucleotide sequences similar to 18S and 28S rRNA of colpodellids, chromerids, other apicomplexans, and related groups, i.e. principal representatives of Myzozoa [14], was conducted in nr/nt and metagenomic wgs databases maintained at NCBI (www.ncbi.nlm.nih.gov) using BLAST [15]. To thoroughly account for diversity, representatives of the most isolated clades were also selected. Partial nucleotide sequences were assembled into contigs in the cases when the similarity of overlapping regions was above 97%. Names and Accession Numbers of selected sequences are presented in Figs. 1 and 2. Considering that the species diversity available for the 28S rRNA genes is lower than for the 18S rRNA, but does not overlap completely, two sets of sequences were generated: one for the 28S rRNA including 50 operational taxonomic units (OTUs) and another one for the 18S rRNA (81 OTUs), the latter supplemented with 28S rRNA when possible. Alignments obtained with the MUSCLE program [16] were corrected manually in the BioEdit editor [17]. Trees were constructed with the MrBayes 3.2.2 program [18] under the GTR model accounting for invariant sites and among-site evolutionary rate heterogeneity with 10 categories. For the concatenated alignment of 18S and 28S rRNA (81 OTUs), all parameters except topology and branch lengths were optimized separately for two partitions (18S and 28S rRNA). For the each set, 5,000,000 Markov Chain Monte Carlo generations were sampled in two independent runs, with a 50% burn-in for tree construction. Convergence was estimated by PSRF values [19]. To evaluate bootstrap support for the nodes of the Bayesian tree, we used the RAxML program [20] that searches for trees using the maximum likelihood method. The following parameters were selected for analysis with RAxML: GTR evolutionary model with four categories of gamma-distributed among-site evolutionary rate heterogeneity and evaluation of the fraction of invariant sites; to calculate support values, 1000 bootstrap replicas were used. Significance of differences of alternative topologies was evaluated using the Bayesian factor [21] (by calculating the logarithm of marginal likelihood with the *ss* command in the MrBayes 3.2.2 program [18]) and by the AU-test (using the CONSEL program [22]). Alternative topologies for the AU-test were constructed by transferring one of the three clades (*Chromera velia*, *Vitrella brassicaformis*, and *Alphamonas edax* with close OTUs if available) of the Bayesian tree in the TreeView editor [23]. The site-wise likelihood values were calculated with TREE-PUZZLE v. 5.2 [24] using *wsl* command; the nucleotide substitution rates for the GTR model were taken from the results of the Bayesian analysis.

RESULTS AND DISCUSSION

The 28S rRNA genes of *Colpodella angusta*, *Voromonas pontica*, and *Alphamonas edax* colpodellids are similar in primary structure to the corresponding genes of sporozoans and dinoflagellates and do not display noticeable deviations in size or location of hypervariable regions. The 28S rRNA of the investigated *V. pontica* strain is a 100% match along the overlapping 1197 bp region with the sequence of the *V. pontica* strain HFCC41 published previously [25]. In addition to the *V. pontica* HFCC41, 10 partial sequences of the 28S rRNA genes from unidentified species belonging to the colpodellid

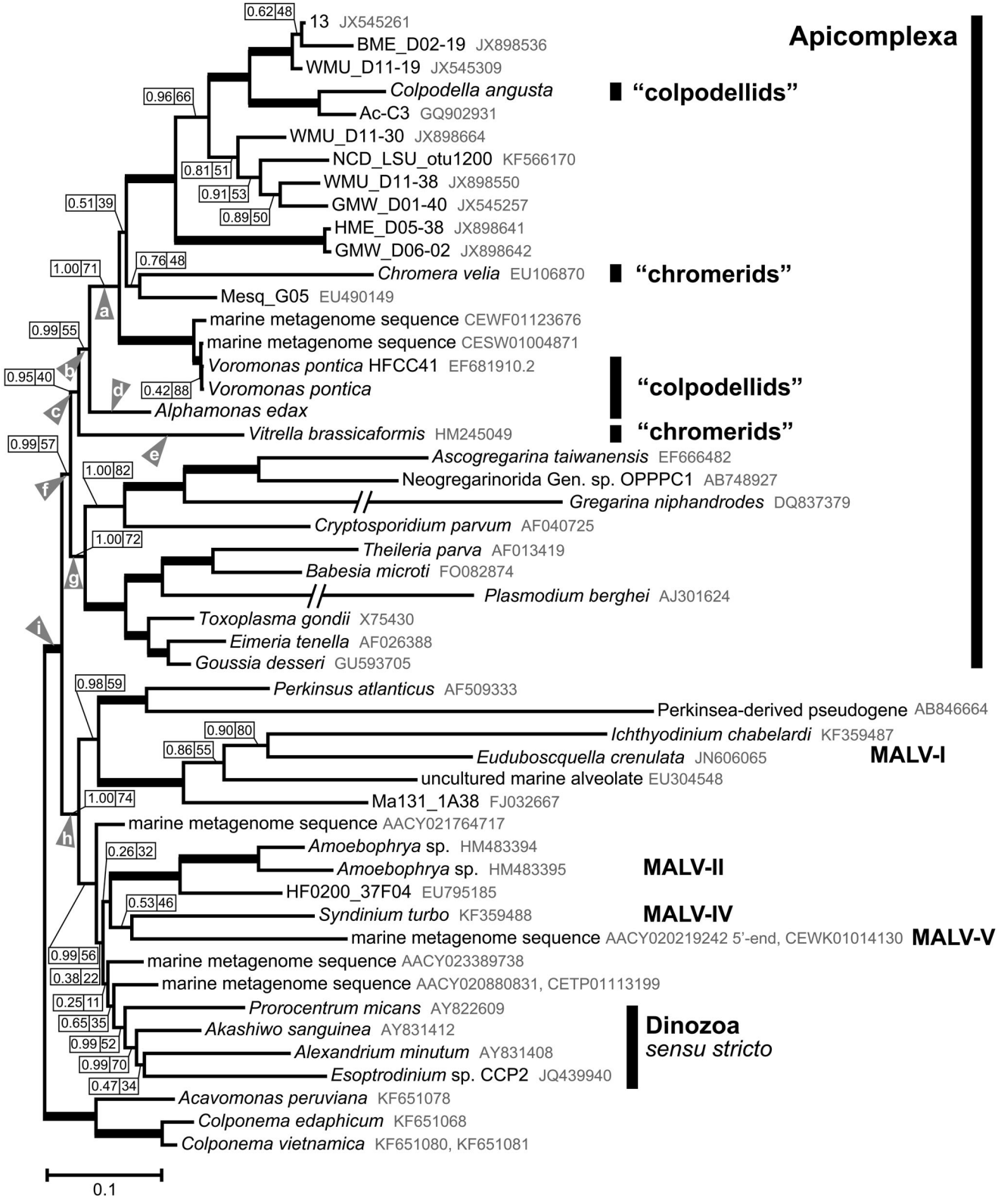

Fig. 1. Bayesian tree of 28S rRNA. Edges with posterior probability and ML bootstrap index above 90% are highlighted in bold; for the remaining branches, the values of posterior probability/bootstrap index are presented. The length of branches to *Gregarina niphandrodes* and *Plasmodium berghei* is shortened by a third. Gray triangles (a-i) represent points of clade transfers for the AU-test (see Table 2 for AU-test *p*-values).

clade and one sequence relatively similar to *Chromera velia* were found in the NCBI nr/nt database (Fig. 1). They were all obtained from soil samples [26, 27] or cave habitats [28]. Additionally, we found contigs with 98-100% identity to 28S rRNA of *V. pontica* in the marine metagenome assemblies [29] (wgs databases). This agrees with the previous observation that *Voromonas* is a marine genus [12]. It can be characterized as the most abundant colpodellid in the ocean plankton because it is the only genus of colpodellids found in the wgs metagenomic databases of ocean plankton.

The Bayesian tree of 28S rRNA (Fig. 1) was constructed after reaching convergence of two independent runs, in accordance with a criterion [19]. The tree unites free-living colpodellids, both species of chromerids, and related unidentified species in one common clade, which was recently named chrompodellids [12], with posterior probability of 0.95. The branch of the autotrophic species *Vitrella brassicaformis* diverges first, followed by the heterotrophic *A. edax*, and the remaining clade includes heterotrophic species *C. angusta* and *V. pontica* and autotrophic species *Chromera velia*. Hence, the heterotrophic and autotrophic species are mixed in the tree and do not form monophyletic groups according to trophism. The 28S rRNA tree topology of chrompodellids is identical to the one obtained through the phylogenetic analysis of concatenated amino acid sequences of 85 proteins reconstructed from transcriptome data [12], and chrompodellids as a whole are a sister group to the rest of parasitic Apicomplexa (Sporozoa). The split of the latter into coccidia/haemosporidia clade and gregarine/*Cryptosporidium* clade corresponds to the commonly accepted notions [30].

The closest outgroup to Apicomplexa comprises other representatives of Myzozoa [1, 15] (designated here as Dinozoa *sensu lato*) including dinoflagellates (i.e. Dinozoa *sensu stricto*), perkinsids [31], and related taxons that contain species with highly uneven rates of molecular evolution. It is, therefore, important to achieve representative sampling for accurate reconstruction of the apicomplexan phylogenetic tree. However, the 28S rRNA genes are available only for a few species in this clade with the exception of dinoflagellates. To achieve better sampling we included the 28S rRNA sequences of *Perkinsus*, a pseudogene of the stramenopile *Ciliophrys infusionum* acquired by horizontal transfer from an unknown likely parasitic perkinsid [32], sequences from the marine alveolate group I (MALV-I) [33-36], and partial sequences of Syndinales: *Syndinium turbo* (MALV-IV) and *Amoebophrya* spp. (MALV-II) [36, 37]. The complete ribosomal operon of an unknown representative of MALV-II is available under the name HF0200_37F04 (Accession Number EU795185 in GenBank). The inclusion of HF0200_37F04 in MALV-II is supported by notable similarity of its 18S and 28S rRNA nucleotide sequences to those of *Amoebophrya* spp. Additionally, another contig assembled from different wgs libraries [29, 38] representing the metagenome of ocean plankton (GenBank Accession Numbers AACY022815131, CEWK01014130, and AACY020219242, 5′-terminus) overlaps by 415 bp (with a single nucleotide replacement) with the 18S rRNA sequence assigned to MALV-V [39]. Based on this observation, we assign the obtained contig to the MALV-V group. Furthermore, several sequences from marine metagenomes occupy an isolated position in the tree, and their relationship to the taxa identified earlier remains to be elucidated (Fig. 1).

The majority of groups, Apicomplexa in particular, enjoy high posterior probability indexes in the Bayesian tree. The bootstrap support of these nodes from the maximum likelihood analysis provides a less confident estimate (Fig. 1). For this reason, we tested statistical significance of the phylogenetic relationships suggested by the Bayesian inference. Comparison of the Bayes factor [21] under the topological constraint of the chrompodellid group monophyly against the forcedly ruled out monophyly strongly rejects the second variant (Table 1). The AU-test rejects any other position of the chromerid *Chromera velia* (under transition along the tree together with the sister OTU Mesq_G05, Accession Number EU490149 in GenBank) except in the clade with *C. angusta* and *V. pontica* (Table 2), and any other position of the *A. edax* except its position in the ML tree or in the monophyletic group with *V. brassicaformis*. At the same time, the AU-test does not reject some other alternatives for the position of *V. brassicaformis* in the tree (Table 2).

Using the 18S rRNA gene fragments from various environmental samples as well as from a few identified species [5, 15, 40, 41], the number of OTUs related to colpodellids and chromerids can be extended, but the diversity is not increased significantly because the majority of OTUs belong to one of the four clades described above [12]. The fifth clade, for which there are no identified species or any gene sequences available so far except 18S rRNA, groups with the clade that includes *Colpodella*, *Chromera*, and *Voromonas* genera. They are united in the combined tree (constructed from the 18S rRNA nucleotide sequences supplemented with the 28S rRNA sequences when possible) with posterior probability of 0.97 (Fig. 2). In general, the topology of the combined tree is in agreement with the topology of the 28S rRNA tree; moreover, the values of posterior probabilities and bootstrap indexes for the same clades are close. The Bayes factor comparison for the combined data and for the 28S rRNA under the condition of topological constraint is strongly in favor of the chrompodellid monophyly (Table 1). The AU-test (Table 2) rejects all violation of the chrompodellid monophyly and alternative positions of its largest clades with the exception of a group uniting the *V. brassicaformis* clade with the *A. edax* clade (probability 0.613) or with parasitic Apicomplexa (probability 0.379).

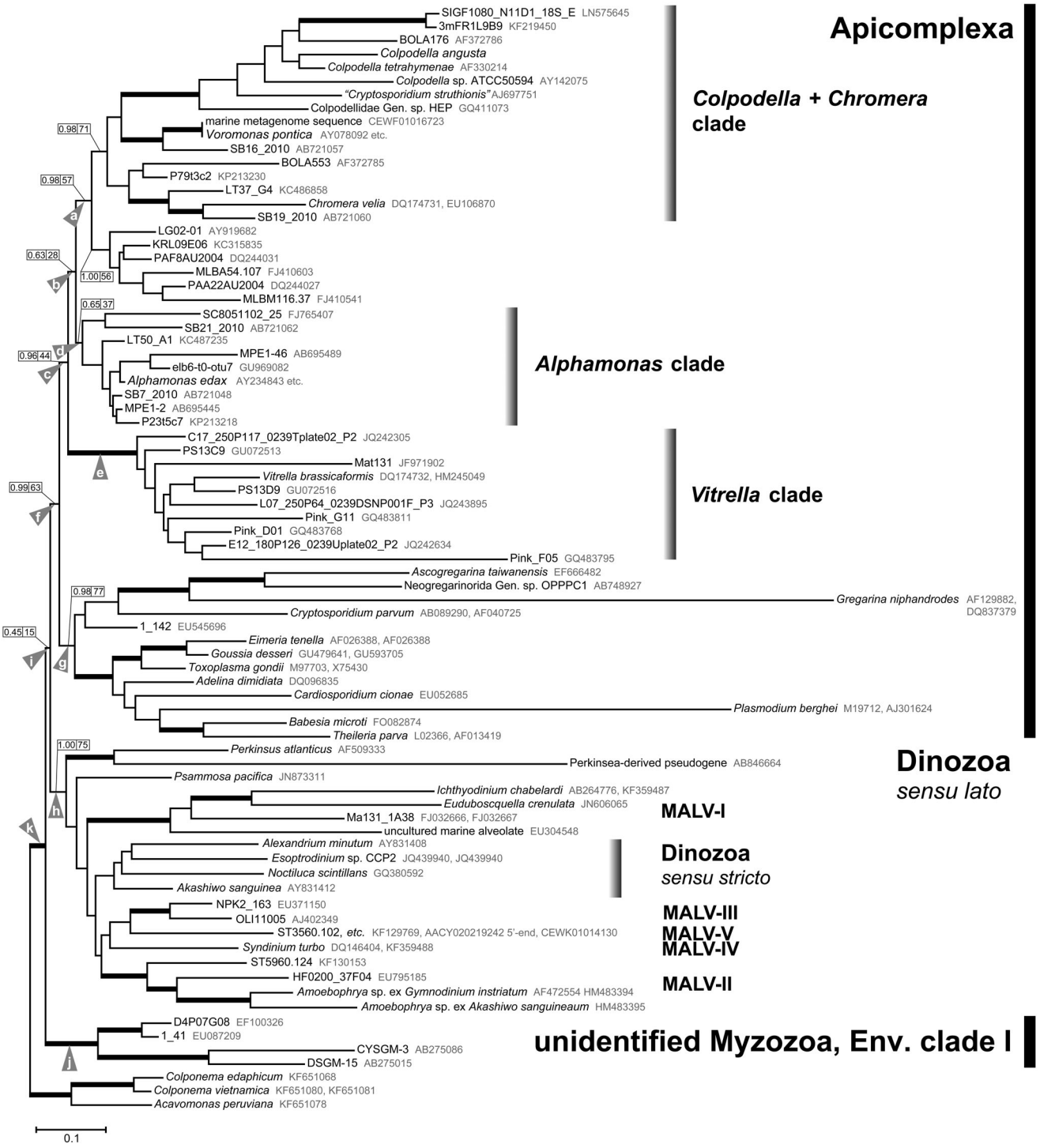

Fig. 2. Bayesian tree constructed from concatenated alignment of 18S rRNA and 28S rRNA; the sequences of 28S rRNA were available only for some OTUs. Edges with posterior probability and ML bootstrap index above 90% are highlighted in bold; for the remaining branches, the values of posterior probability/bootstrap index are presented. Gray triangles (a-k) represent points of clade transfers for the AU-test (see Table 2 for AU-test *p*-values).

While for Apicomplexa, the topologies of the 28S rRNA and the combined 18S and 28S rRNA trees agree and the statistical support values are similar, the perkinsids/dinoflagellates clade displays differences in topologies. It is likely that the differences are due to the fuller taxonomic set of the 18S rRNA, in which the nucleotide sequences of *Psammosa* [42] and marine alveolates MALV-III are present. In addition, the set of 18S rRNA

gene fragments is extended by the non-classified Myzozoa, which were recently designated as "clade I" [12]. This group does not contain many species and is equidistant from Apicomplexa and Dinozoa *sensu lato*. Genes of its representatives have been found in oxygen-depleted subtidal [43] and deep-sea marine sediments [44, 45] or in enriched culture of deep-sea methane cold seep obtained under anaerobic conditions [45]. One of the phylotypes (CYSGM-3, Accession Number AB275086) was found only in the enriched culture, but not in the initial sample, which could be explained by its reproduction under conditions of heterotrophic anaerobic culture. According to phylogenetic trees published previously, "clade I" was located at the base of Dinozoa *sensu lato* with low statistical support [11, 12, 45]. When alternative topologies were tested for 81 OTUs alignments, the AU-test estimated the probabilities of uniting "clade I" with Dinozoa *sensu lato*, apicomplexans, sporozoans, or chrompodellids as 0.747, 0.746, 0.15, and 0.013, respectively. The AU probability of topology that places "clade I" at the root of Myzozoa (Fig. 2) is 0.747. Hence, the test assigns equal probability to the three possible variants of relative positions of the major Myzozoa clades: Apicomplexa, Dinozoa *sensu lato*, and "clade I". It is of great interest to obtain more detailed data on the morphology, life cycles, and genetics of the species from "clade I".

The previous hypotheses on the phylogenetic relationships of colpodellids and chromerids were based for the most part on taxonomically poor sets or sets limited to 18S rRNA genes, and they were in rather poor agreement. Among those hypotheses, the following were proposed: monophyly of colpodellids based on a sample not including chromerids [41]; monophyly of *Colpodella* with Apicomplexa, and of *Voromonas* and *Alphamonas* – with Dinozoa *sensu lato* [15]; colpodellids and chromerids form mixed clades (*Colpodella* + *Voromonas* + *Chromera*) and (*Alphamonas* + *Vitrella*), which are monophyletic [10] or paraphyletic relative to sporozoans [9], or form with them unresolved trichotomy [11]; multiple paraphyly of colpodellids (Fig. 1 in [12]). The phylogenetic analysis of the extended taxonomic set with 28S rRNA gene sequences presented in this work (Fig. 1), as well as in its combination with 18S rRNA (Fig. 2), groups photosynthetic algae chromerids (*Chromera* and *Vitrella*) and predatory colpodellids (*Colpodella*, *Voromonas*, *Alphamonas*) into a monophyletic group with the parasitic Apicomplexa (sporozoans) as closest relatives. The obtained tree topology for the major clades is in complete agreement with the tree reconstructed from an alignment of 85 proteins [12], differing only by the lower support of the edge between the "*Vitrella brassicaformis*" and "*Alphamonas edax*" clades: the AU-test for rRNA does not reject the monophyly of the indicated clades [9-11], however, the likelihood of this topology is not maximal. The synapomorphies of *A. edax*, *C. angusta*, and *V. pontica* also support the maximum likelihood tree of the rRNA [12]: transfer of the *sufB* gene with an unknown function in biogenesis of FeS clusters into a nucleus; in *V. brassicaformis*, the *sufB* gene retains ancestral localization in the plastid genome. The agreement of the phylogenetic tree based on the traditional rRNA with the phylogenom-

Table 1. Logarithm of marginal likelihood under topological constraints

Type of data	Topological constraint	
	monophyly of chrompodellid accepted	monophyly of chrompodellid rejected
28S rRNA	–38350.98	–38384.73
18S and 28S rRNA	–63538.05	–63596.79

Table 2. Probability of alternative topologies (AU test)

Point of transfer	Moving clade (name according to identified species)		
	"*Chromera velia*"	"*Alphamonas edax*"	"*Vitrella brassicaformis*"
	28S rRNA tree		
a	0.643	–	0.018*
b	0.018*	–	–
c	0.004*	0.018*	–
d	0.003*	–	0.276
e	0.044*	0.276	–
f	3e-009*	0.006*	0.020*
g	8e-008*	0.008*	0.470
h	1e-006*	0.003*	0.008*
i	1e-005*	0.002*	0.224
	Tree for combined 18S and 28S rRNA		
a	0.035*	–	0.019*
b	2e-004*	–	–
c	0.001*	0.019*	–
d	1e-007*	–	0.613
e	2e-004*	0.613	–
f	2e-004*	0.003*	0.002*
g	3e-004*	5e-008*	0.379
h	6e-005*	6e-072*	3e-005*
i	9e-005*	7e-029*	0.019*
j	8e-006*	9e-097*	0.019*
k	1e-031*	8e-094*	0.020*

Note: The points of transfer for the 28S rRNA and concatenated gene trees are indicated in Figs. 1 and 2.
* Difference in likelihood of this topology versus the Bayesian inference topology is statistically significant.

ic one eliminates previous contradictions and consolidates the current view on the phylogeny of colpodellids and chromerids.

This study was supported by the Russian Science Foundation (grants No. 14-50-00029, phylogenetic analysis; No. 14-14-00515, determination of nucleotide sequences of *C. angusta*), the Russian Foundation for Basic Research (grant Nos. 14-04-00553, 15-34-20065, 15-29-02518, determination of nucleotide sequences of other species), the Program of the President of Russian Federation (grant No. MK-7436.2015.4), and the European Center of Ichtyoparasitology (grant No. GBP505/12/G112, work with databases).

Calculations were conducted using the supercomputer at Lomonosov Moscow State University.

REFERENCES

1. Gajadhar, A. A., Marquardt, W. C., Hall, R., Gunderson, J., Ariztia-Carmona, E. V., and Sogin, M. L. (1991) Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata*, and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates, *Biochem. Parasitol.*, **45**, 147-154.
2. Mukhina, V. S. (2014) Origination and evolution of plastids, *Zh. Obshch. Biol.*, **75**, 329-352.
3. McFadden, G. I., Reith, M. E., Munholland, J., and Lang-Unnasch, N. (1996) Plastid in human parasites, *Nature*, **381**, 482.
4. Krylov, M. V., and Myl'nikov, A. P. (1986) New taxons in the type Sporozoa, Spiromonadomorphina subcl. n. Spiromonadida ordo n., *Parazitologiya*, **20**, 425-430.
5. Kuvardina, O. N., Leander, B. S., Aleshin, V. V., Myl'nikov, A. P., Keeling, P. J., and Simdyanov, T. G. (2002) The phylogeny of colpodellids (Alveolata) using small subunit rRNA gene sequences suggests they are the free-living sister group to apicomplexans, *J. Eukaryot. Microbiol.*, **49**, 498-504.
6. Leander, B. S., and Keeling, P. J. (2003) Morphostasis in alveolate evolution, *Trends Ecol. Evol.*, **18**, 395-402.
7. Moore, R. B., Obornik, M., Janouškovec, J., Chrudimsky, T., Vancova, M., Green, D. H., Wright, S. W., Davies, N. W., Bolch, C. J., Heimann, K., Slapeta, J., Hoegh-Guldberg, O., Logsdon, J. M., and Carter, D. A. (2008) A photosynthetic alveolate closely related to apicomplexan parasites, *Nature*, **451**, 959-963.
8. Obornik, M., Modry, D., Lukes, M., Cernotikova-Stribrna, E., Cihlar, J., Tesarova, M., Kotabova, E., Vancova, M., Prasil, O., and Lukes, J. (2012) Morphology, ultrastructure and life cycle of *Vitrella brassicaformis* n. sp., n. gen., a novel chromerid from the Great Barrier Reef, *Protist*, **163**, 306-323.
9. Cavalier-Smith, T. (2014) Gregarine site-heterogeneous 18S rDNA trees, revision of gregarine higher classification, and the evolutionary diversification of Sporozoa, *Eur. J. Protistol.*, **50**, 472-495.
10. Gile, G. H., and Slamovits, C. H. (2014) Transcriptomic analysis reveals evidence for a cryptic plastid in the colpodellid *Voromonas pontica*, a close relative of chromerids and apicomplexan parasites, *PLoS One*, **9**, e96258.
11. Mikhailov, K. V., Janouškovec, J., Tikhonenkov, D. V., Mirzaeva, G. S., Diakin, A. Y., Simdyanov, T. G., Mylnikov, A. P., Keeling, P. J., and Aleoshin, V. V. (2014) A complex distribution of elongation family GTPases EF1A and EFL in basal alveolate lineages, *Genome Biol. Evol.*, **6**, 2361-2367.
12. Janouškovec, J., Tikhonenkov, D. V., Burki, F., Howe, A. T., Kolisko, M., Mylnikov, A. P., and Keeling, P. J. (2015) Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives, *Proc. Natl. Acad. Sci. USA*, **112**, 10200-10207.
13. Van der Auwera, G., Chapelle, S., and De Wachter, R. (1994) Structure of the large ribosomal subunit RNA of *Phytophthora megasperma*, and phylogeny of the oomycetes, *FEBS Lett.*, **338**, 133-136.
14. Cavalier-Smith, T., and Chao, E. E. (2004) Protalveolate phylogeny and systematics and the origins of Sporozoa and dinoflagellates (phylum Myzozoa nom. nov.), *Eur. J. Protistol.*, **40**, 185-212.
15. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, **25**, 3389-33402.
16. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.*, **32**, 1792-1797.
17. Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucleic Acids Symp. Ser.*, **41**, 95-98.
18. Ronquist, F., Teslenko, M., Van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space, *Syst. Biol.*, **61**, 539-542.
19. Gelman, A., and Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences, *Stat. Sci.*, **7**, 457-472.
20. Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics*, **30**, 1312-1313.
21. Kass, R. E., and Raftery, A. E. (1995) Bayes factors, *J. Am. Stat. Assoc.*, **90**, 773-795.
22. Shimodaira, H., and Hasegawa, M. (2001) CONSEL: for assessing the confidence of phylogenetic tree selection, *Bioinformatics*, **17**, 1246-1247.
23. Page, R. D. (1996) TREEVIEW: an application to display phylogenetic trees on personal computers, *Comput. Appl. Biosci.*, **12**, 357-358.
24. Schmidt, H. A., Strimmer, K., Vingron, M., and Von Haeseler, A. (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing, *Bioinformatics*, **18**, 502-504.
25. Wylezich, C., Nies, G., Mylnikov, A. P., Tautz, D., and Arndt, H. (2010) An evaluation of the use of the LSU rRNA D1-D5 domain for DNA-based taxonomy of eukaryotic protists, *Protist*, **161**, 342-352.
26. Beloqui, A., Nechitaylo, T. Y., Lopez-Cortes, N., Ghazi, A., Guazzaroni, M. E., Polaina, J., Strittmatter, A. W., Reva, O., Waliczek, A., Yakimov, M. M., Golyshina, O. V., Ferrer, M., and Golyshin, P. N. (2010) Diversity of glycosyl

hydrolases from cellulose-depleting communities enriched from casts of two earthworm species, *Appl. Environ. Microbiol.*, **76**, 5934-5946.
27. Hollister, E. B., Schadt, C. W., Palumbo, A. V., Ansley, R. J., and Boutton, T. W. (2010) Structural and functional diversity of soil bacterial and fungal communities following woody plant encroachment in the southern Great Plains, *Soil Biol. Biochem.*, **42**, 1816-1824.
28. Zhang, T., Victor, T. R., Rajkumar, S. S., Li, X., Okoniewski, J. C., Hicks, A. C., Davis, A. D., Broussard, K., LaDeau, S. L., Chaturvedi, S., and Chaturvedi, V. (2014) Mycobiome of the bat white nose syndrome affected caves and mines reveals diversity of fungi and local adaptation by the fungal pathogen *Pseudogymnoascus* (*Geomyces*) *destructans*, *PLoS One*, **9**, e116149.
29. De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J. M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horak, A., Jaillon, O., Lima-Mendez, G., Lukes, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Tara Oceans Coordinators – Acinas, S. G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M. E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P., and Karsenti, E. (2015) Eukaryotic plankton diversity in the sunlit ocean, *Science*, **348**, doi: 10.1126/science.1261605.
30. Carreno, R. A., Martin, D. S., and Barta, J. R. (1999) *Cryptosporidium* is more closely related to the gregarines than to coccidia as shown by phylogenetic analysis of apicomplexan parasites inferred using small-subunit ribosomal RNA gene sequences, *Parasitol. Res.*, **85**, 899-904.
31. Siddall, M. E., Reece, K. S., Graves, J. E., and Burreson, E. M. (1997) Total evidence refutes the inclusion of *Perkinsus* species in the phylum Apicomplexa, *Parasitology*, **115**, 165-176.
32. Yabuki, A., Toyofuku, T., and Takishita, K. (2014) Lateral transfer of eukaryotic ribosomal RNA genes: an emerging concern for molecular ecology of microbial eukaryotes, *ISME J.*, **8**, 1544-1547.
33. Massana, R., Karniol, B., Pommier, T., Bodaker, I., and Beja, O. (2008) Metagenomic retrieval of a ribosomal DNA repeat array from an uncultured marine alveolate, *Environ. Microbiol.*, **10**, 1335-1343.
34. Marande, W., Lopez-Garcia, P., and Moreira, D. (2009) Eukaryotic diversity and phylogeny using small- and large-subunit ribosomal RNA genes from environmental samples, *Environ. Microbiol.*, **11**, 3179-3188.
35. Coats, D. W., Bachvaroff, T. R., and Delwiche, C. F. (2012) Revision of the family Duboscquellidae with description of *Eduboscquella crenulata* n. gen., n. sp. (Dinoflagellata, Syndinea), an intracellular parasite of the ciliate *Favella panamensis* Kofoid & Campbell, *J. Eukaryot. Microbiol.*, **59**, 1-11.
36. Lepelletier, F., Karpov, S. A., Le Panse, S., Bigeard, E., Skovgaard, A., Jeanthon, C., and Guillou, L. (2014) *Parvilucifera rostrata* sp. nov. (Perkinsozoa), a novel parasitoid that infects planktonic dinoflagellates, *Protist*, **165**, 31-49.
37. Coats, D. W., Kim, S., Bachvaroff, T. R., Handy, S. M., and Delwiche, C. F. (2010) *Tintinnophagus acutus* n.g., n.sp. (phylum Dinoflagellata), an ectoparasite of the ciliate *Tintinnopsis cylindrica* Daday 1887, and its relationship to *Duboscquodinium collini* Grasse 1952, *J. Eukaryot. Microbiol.*, **57**, 468-482.
38. Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H., and Smith, H. O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea, *Science*, **304**, 66-74.
39. Wu, W., Huang, B., Liao, Y., and Sun, P. (2014) Picoeukaryotic diversity and distribution in the subtropical-tropical South China Sea, *FEMS Microbiol. Ecol.*, **89**, 563-579.
40. Siddall, M. E., Reece, K. S., Nerad, T. A., and Burreson, E. M. (2001) Molecular determination of the phylogenetic position of a species in the genus *Colpodella* (Alveolata), *Am. Mus. Novitates*, **3314**, 1-10.
41. Leander, B. S., Kuvardina, O. N., Aleshin, V. V., Mylnikov, A. P., and Keeling, P. J. (2003) Molecular phylogeny and surface morphology of *Colpodella edax* (Alveolata): insights into the phagotrophic ancestry of apicomplexans, *J. Eukaryot. Microbiol.*, **50**, 334-340.
42. Okamoto, N., Horak, A., and Keeling, P. J. (2012) Description of two species of early branching dinoflagellates, *Psammosa pacifica* n.g., n.sp. and *P. atlantica* n.sp., *PLoS One*, **7**, e34900.
43. Stoeck, T., Kasper, J., Bunge, J., Leslin, C., Ilyin, V., and Epstein, S. (2007) Protistan diversity in the Arctic: a case of paleoclimate shaping modern biodiversity? *PLoS One*, **2**, e728.
44. Park, S. J., Park, B. J., Pham, V. H., Yoon, D. N., Kim, S. K., and Rhee, S. K. (2008) Microeukaryotic diversity in marine environments, an analysis of surface layer sediments from the East Sea, *J. Microbiol.*, **46**, 244-249.
45. Takishita, K., Yubuki, N., Kakizoe, N., Inagaki, Y., and Maruyama, T. (2007) Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from raw sediment samples and two enrichment cultures, *Extremophiles*, **11**, 563-576.