Research article

Cyanobacterial contribution to the genomes of the plastid-lacking protists

Shinichiro Maruyama*[1], Motomichi Matsuzaki[1,2,3], Kazuharu Misawa[1,3,3] and Hisayoshi Nozaki[1]

Address: [1]Department of Biological Sciences, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan, [2]Current address: Department of Biomedical Chemistry, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan and [3]Current address: Research Program for Computational Science, Riken, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

Email: Shinichiro Maruyama* - maruyama@biol.s.u-tokyo.ac.jp; Motomichi Matsuzaki - mzaki@m.u-tokyo.ac.jp; Kazuharu Misawa - kazumisawa@riken.jp; Hisayoshi Nozaki - nozaki@biol.s.u-tokyo.ac.jp

* Corresponding author

Abstract

Background: Eukaryotic genes with cyanobacterial ancestry in plastid-lacking protists have been regarded as important evolutionary markers implicating the presence of plastids in the early evolution of eukaryotes. Although recent genomic surveys demonstrated the presence of cyanobacterial and algal ancestry genes in the genomes of plastid-lacking protists, comparative analyses on the origin and distribution of those genes are still limited.

Results: We identified 12 gene families with cyanobacterial ancestry in the genomes of a taxonomically wide range of plastid-lacking eukaryotes (*Phytophthora* [Chromalveolata], *Naegleria* [Excavata], *Dictyostelium* [Amoebozoa], *Saccharomyces* and *Monosiga* [Opisthokonta]) using a novel phylogenetic pipeline. The eukaryotic gene clades with cyanobacterial ancestry were mostly composed of genes from bikonts (Archaeplastida, Chromalveolata, Rhizaria and Excavata). We failed to find genes with cyanobacterial ancestry in *Saccharomyces* and *Dictyostelium*, except for a photorespiratory enzyme conserved among fungi. Meanwhile, we found several *Monosiga* genes with cyanobacterial ancestry, which were unrelated to other Opisthokonta genes.

Conclusion: Our data demonstrate that a considerable number of genes with cyanobacterial ancestry have contributed to the genome composition of the plastid-lacking protists, especially bikonts. The origins of those genes might be due to lateral gene transfer events, or an ancient primary or secondary endosymbiosis before the diversification of bikonts. Our data also show that all genes identified in this study constitute multi-gene families with punctate distribution among eukaryotes, suggesting that the transferred genes could have survived through rounds of gene family expansion and differential reduction.

Background

Cyanobacterial ancestors gave rise to plastids (chloroplasts) in the ancestor of a eukaryotic lineage. The birth of the plastid had an impact on eukaryotic genome evolution, by way of endosymbiotic gene transfer (EGT), a particular form of lateral gene transfer (LGT) from endosymbionts into the phylogenetically discontiguous host genome [1]. Subsequently, an algal ancestor gave rise

to secondary plastids in several punctate lineages of eukaryotes. A number of these secondarily phototrophic lineages lost their photosynthetic ability and further diverged into secondarily heterotrophic, plastid-lacking protists [2,3].

Although the position of the root of eukaryotes is still uncertain, the presence of gene fusions and insertion/deletion sequences in the marker genes have allowed us to sort eukaryotes into at least three large groups; Opisthokonta, Amoebozoa and bikonts (Archaeplastida, Chromalveolata, Rhizaria and Excavata) [4-10] (Figure 1). Most phototrophic eukaryotes harboring plastids derived from primary endosymbiosis (primary plastids) are classified into the super-group Archaeplastida (i.e. glauco-

phytes, green plants and red algae) [10]. Although it is widely accepted that primary plastids share a single origin [[11-13], but see [14,15]] and the Archaeplastida are monophyletic [[3,16], but see [17,18]], the evolutionary history of the primary plastids is still debatable [19-21]. In plastid-lacking protists, 'plastid imprints' can be exemplified by genomic information, i.e. genes with affinity to extant cyanobacterial or algal genes. These genes were supposed to have originated from EGT events, and this assumption should be affirmed by the resulting phylogenetic relationship between 'imprint' genes and the extant relatives of the putative endosymbionts. The biggest challenge and the limitation of this 'imprint' searching process is that the inevitable incompleteness of genome information on lineages of interest and the ever-developing phyl-
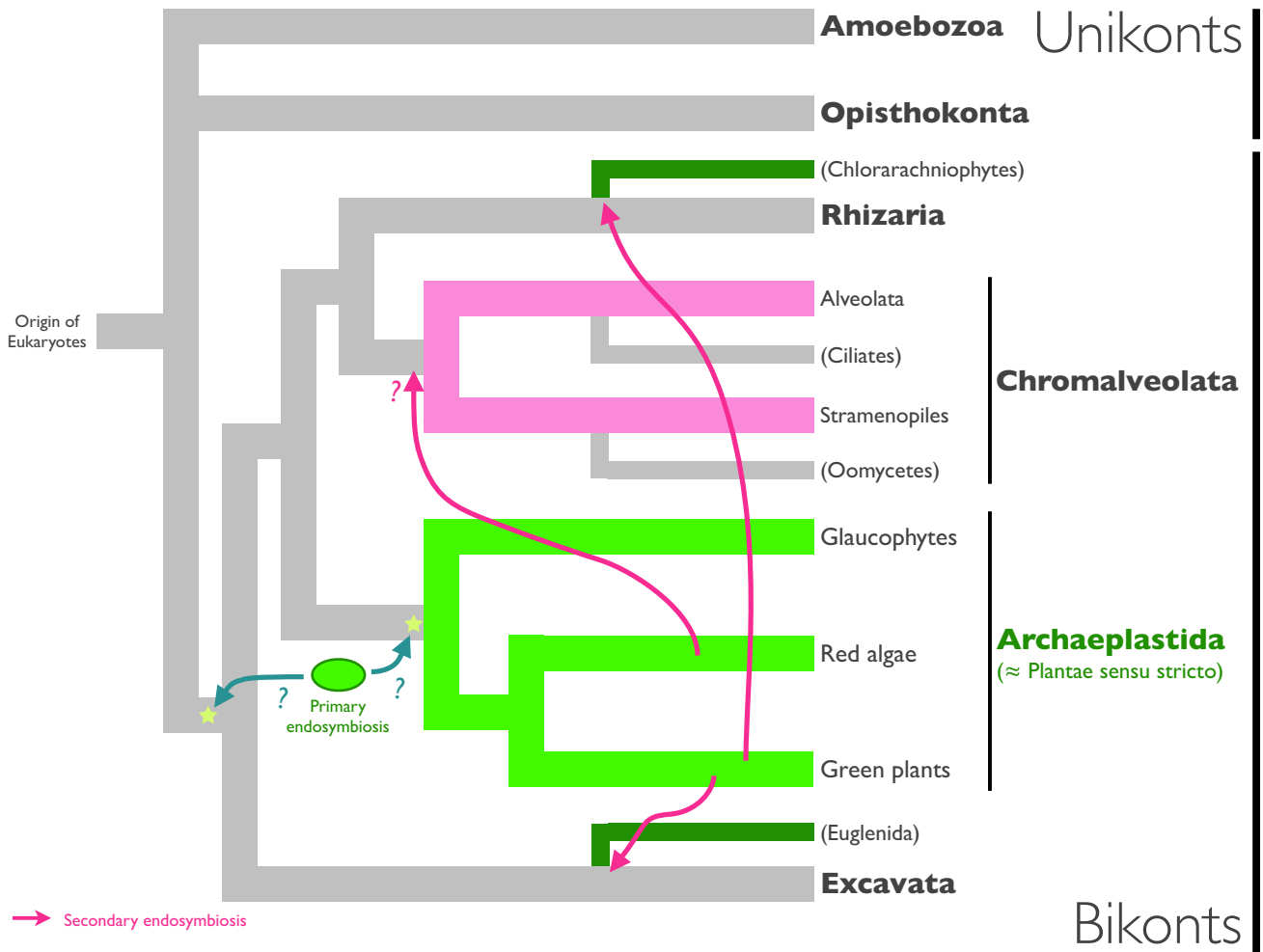
Figure 1
A schematic representation of eukaryotic phylogeny. The current consensus phylogeny and rooting of eukaryotes based on previous studies [4,23,40]. Arrows and stars indicate plastid acquisition via endosymbiosis and alternative hypothetical time points which primary endosymbiosis occurred, respectively. The root of the eukaryotic tree on the unikonts/bikonts boundary is hypothesized, but still controversial [5,7-9,19]. Archaeplastida are represented as a monophyletic group, but see also [19].

ogenetic methodologies make it difficult to distinguish EGT and ancient LGT [22]. Thus, although available eukaryotic genome data are increasingly accumulating, gene and genome phylogenies should be carefully interpreted to infer evolutionary scenarios.

Chromalveolata is a large taxonomic group of eukaryotes, encompassing secondary phototrophs and secondarily heterotrophic protists [10], and the 'chromalveolate hypothesis' argues that this group originated from a common ancestor harboring the chlorophyll *c*-containing secondary plastid derived from a red alga (Figure 1) [23]. Among the secondarily heterotrophic chromalveolates, several lineages have retained remnant chloroplasts for non-photosynthetic metabolic pathways, e.g. apicoplasts in apicomplexan parasites [24]. Recent genomic surveys revealed the presence of plastid-derived genes, and further suggested the presence of cryptic secondary plastids in non-photosynthetic alveolate protists [25,26]. Furthermore, re-examination of the whole genome sequences suggested the existence of algal genes in ciliates, another plastid-lacking alveolate lineage, which could support the photosynthetic ancestry of ciliates [27]. Oomycetes are plastid-lacking stramenopiles, or chromists, classified into Chromalveolata [10]. Although whole genome sequence analysis showed that a number of genes with affinity to photosynthetic organisms (cyanobacteria and algae) are encoded in the nuclear genome, most of these 'plastid imprints' candidates were only suggested by similarity search and phylogenetic analyses have not yet led to fully recovering the expected tree topology [28]. Considering the uncertain phylogenetic affinity of the 'best hit' in similarity search [29], reassessment of the genome information is important to determine whether the evolutionary history of oomycetes is comparable to ciliates [27].

One candidate of 'plastid imprints' in oomycetes has been confirmed by studies reporting the phylogeny of *gnd* genes, which encode 6-phosphogluconate dehydrogenase, showing that some plastid-lacking protists have plant-like, cyanobacterium-derived *gnd* genes [20,21,30]. These analyses suggested that the *gnd* genes with cyanobacterial ancestry were acquired early in eukaryotic evolution, either via ancient eukaryote-to-eukaryote LGT, or primary EGT that occurred earlier than had ever been thought [21]. Additionally, the phylogeny of *gnd* genes demonstrated that cyanobacterial genes are also present in several Excavata protists, e.g. the heterolobosean amoeboflagellate *Naegleria gruberi*. *Naegleria gruberi* is a non-parasitic heterotrophic species related to *N. fowleri*, which is the causative agent of primary amoebic meningoencephalitis in mammals [31]. Although the phylogenetic relationship within Excavata is still unclear, Heterolobosea, together with Jakobida, is likely to be a sister group of Euglenozoa [18,32].

To address how many genes have cyanobacterial ancestry in plastid-lacking protists, and whether cyanobacterial ancestry is limited to this *gnd* gene or also found in other genes, we conducted a phylogenomic analysis using genome sequence data of a taxonomically wide range of plastid-lacking eukaryotes. Here we present a gene mining study with a novel pipeline automatically producing and summarizing one-by-one phylogenetic trees, and show phylogenetic analyses of resultant candidate genes with cyanobacterial ancestry, using the whole genome sequence data from a wide range of eukaryotic lineages.

Results

To address how many genes are derived from cyanobacteria in non-photosynthetic protists, we conducted cyanobacterial gene mining using the genome sequence data of a wide range of the plastid-lacking eukaryotes (Additional file 1). Using the whole genome data, we conducted BLAST searches against all 'Bacteria' and selected queries showing the highest similarity to genes in the available cyanobacteria genome sequences. We then drew the neighbor-joining (NJ) trees for genes showing homology to cyanobacterial counterparts. After the first tree construction step, we selected the gene trees where cyanobacteria and eukaryotes formed a monophyletic group excluding other prokaryotes. As a result, we obtained a shorter list of candidates, which we termed 'genes with cyanobacterial affinity'. Subsequently we re-analyzed the eukaryotic genes with cyanobacterial affinity by visually checking and re-drawing the Bayesian and maximum likelihood (ML) trees after manually trimming operational taxonomic units (OTUs). In total, we identified 12 plastid-lacking protist genes 'with cyanobacterial ancestry' in the genomes of the wide range of eukaryotes: two plastid-lacking bikonts (the oomycete *P. ramorum* and the heterolobosean *N. gruberi*) and three unikonts (the slime mold *D. discoideum*, the budding yeast *S. cerevisiae* and the choanoflagellate *M. brevicollis*) (Table 1). These were the eukaryotic genes with cyanobacterial ancestry that shared the same origin with Archaeplastida and other eukaryotes. They were placed within a monophyletic subclade mostly composed of photosynthetic organisms (cyanobacteria and plants/algae) and showed an apparent cyanobacterial ancestry as far as was determined by tree topology (Table 1; Figures 2, 3, 4 and 5; and Additional files 2, 3, 4, 5, 6, 7, 8 and 9). We found another type of gene with cyanobacterial ancestry, which were the protist genes forming monophyletic groups mostly with genes from extant cyanobacteria (prokaryote-type genes with cyanobacterial ancestry). Among a number of candidate genes found through the first screening, we have presented three typical trees that were resolved with significant support values (Additional files 10, 11 and 12). We postulate that these prokaryote-type genes are remnants of the bacterium-to-eukaryote LGT, which occurred 'recently' in evolution.
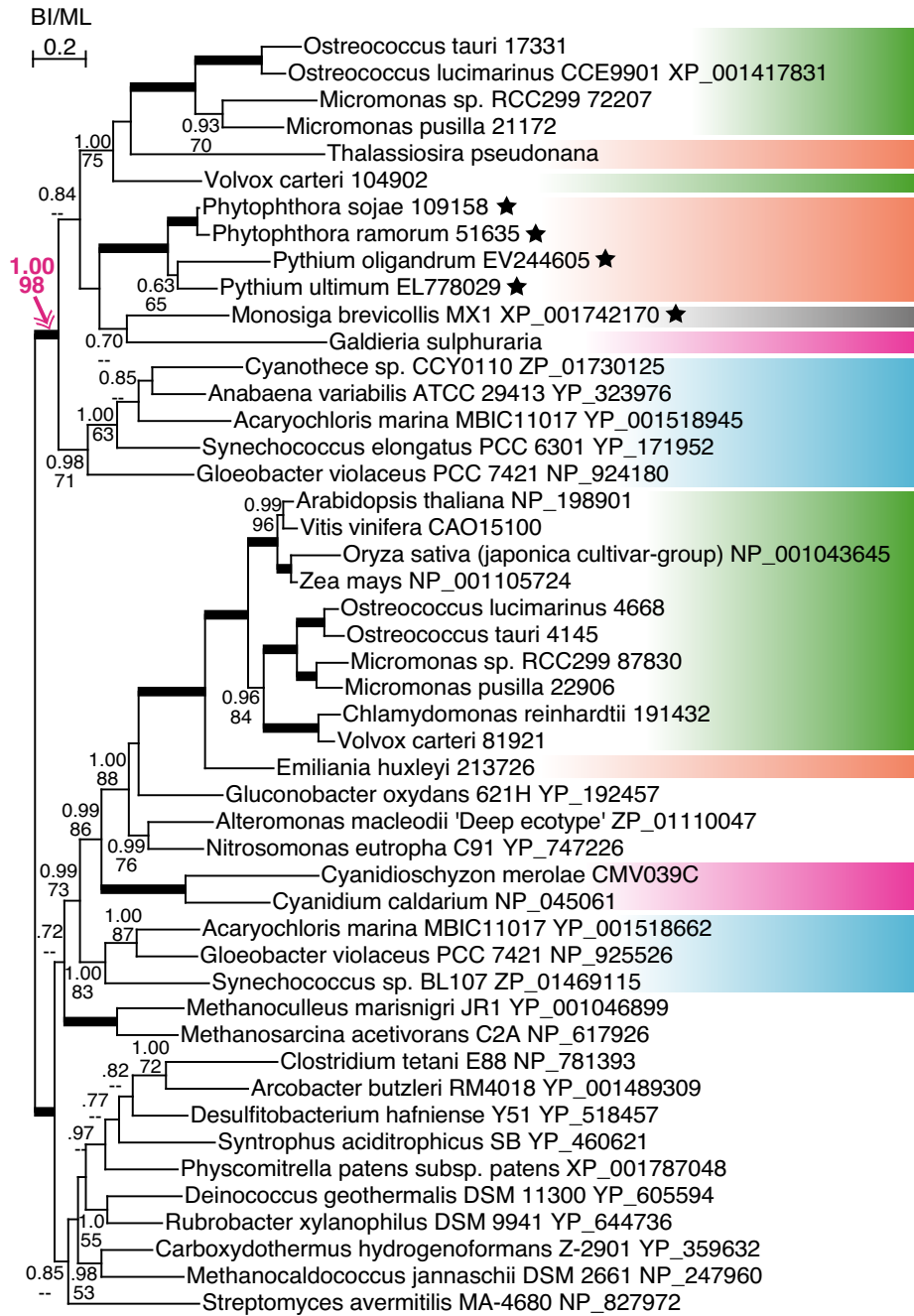
Figure 2
Uroporphyrin III methyltransferase gene phylogeny showing the presence of genes with cyanobacterial ancestry in oomycetes. The MrBayes consensus tree with Bayesian posterior probabilities (BI) (70% or more) and maximum likelihood (ML) bootstrap support values (50% or more) is shown. Thick branches represent BI and ML values not lower than 100 and 95, respectively. Different phylogenetic affiliations are represented as follows: green, green plants; magenta, red algae; blue-green, glaucophytes; orange, Chromalveolata; dark blue, Excavata; yellow, Rhizaria; gray, unikonts; sky blue, cyanobacteria. Stars indicate plastid-lacking eukaryotes.

Figure 3
Cobalamin-independent methionine synthase genes in oomycetes are monophyletic with algal and cyanobacterial homologs. See legend for figure 2.

Figure 4
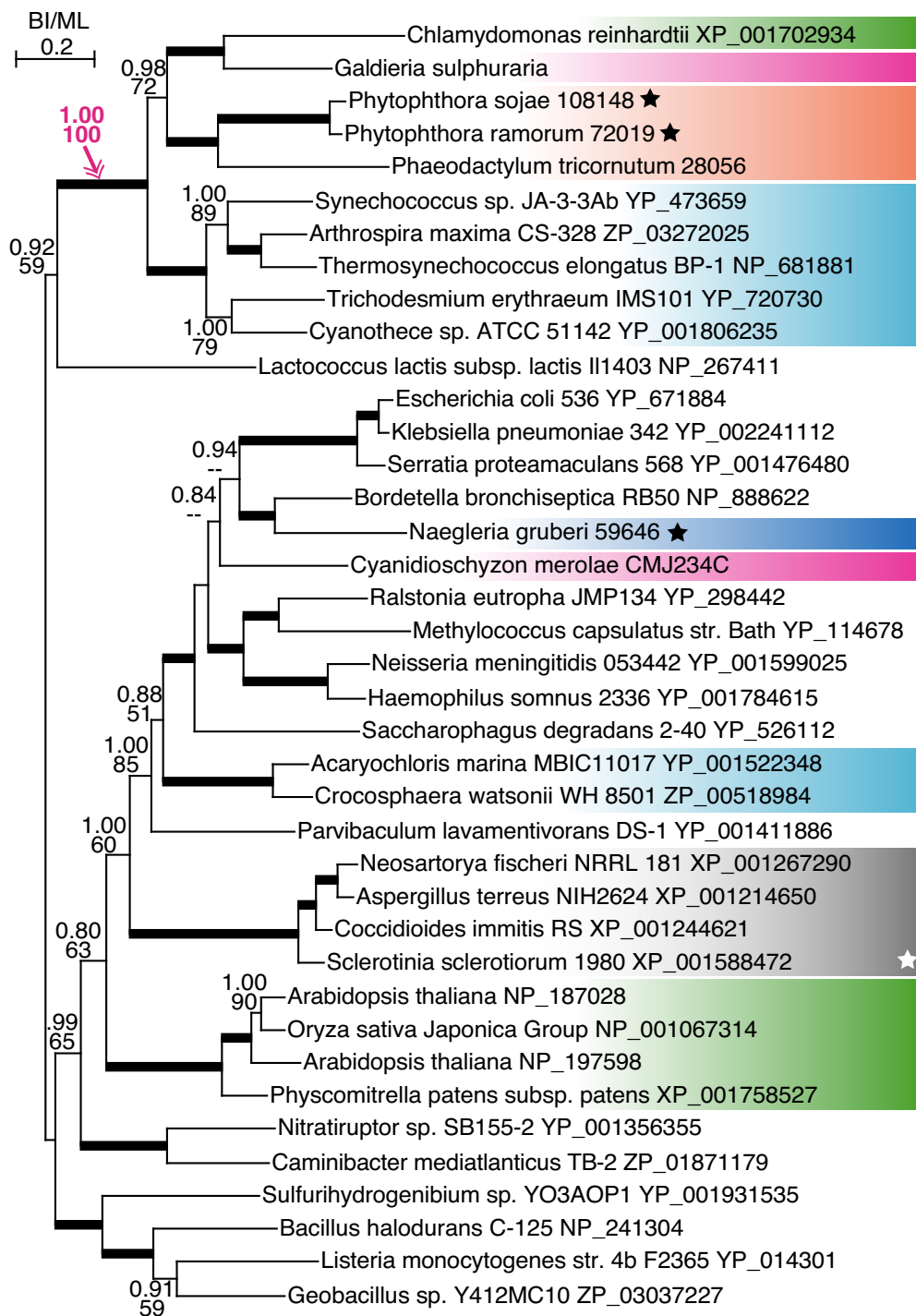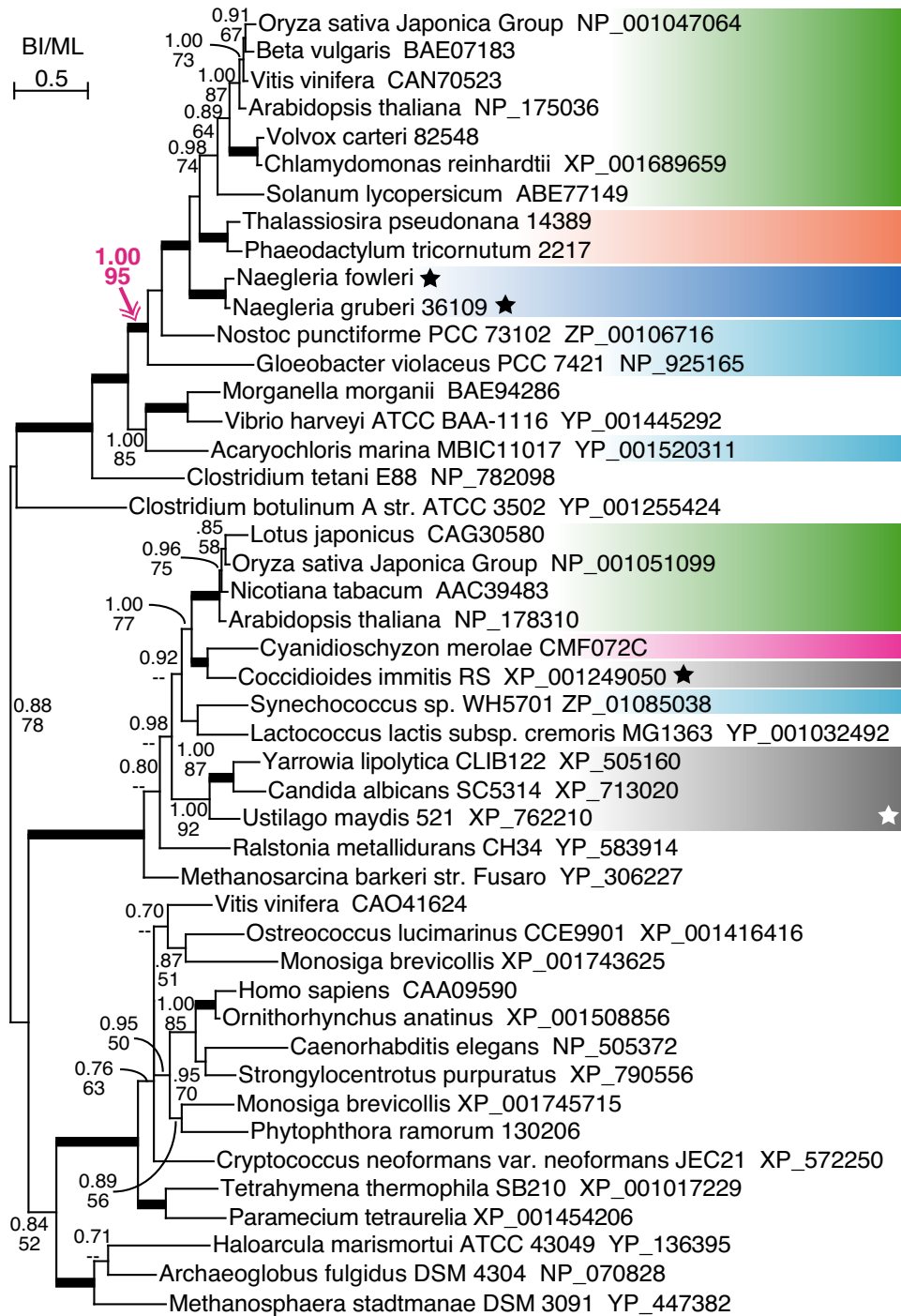**Amino acid decarboxylase genes in Heterolobosea, within a subfamily with cyanobacterial ancestry**. See legend for figure 2.
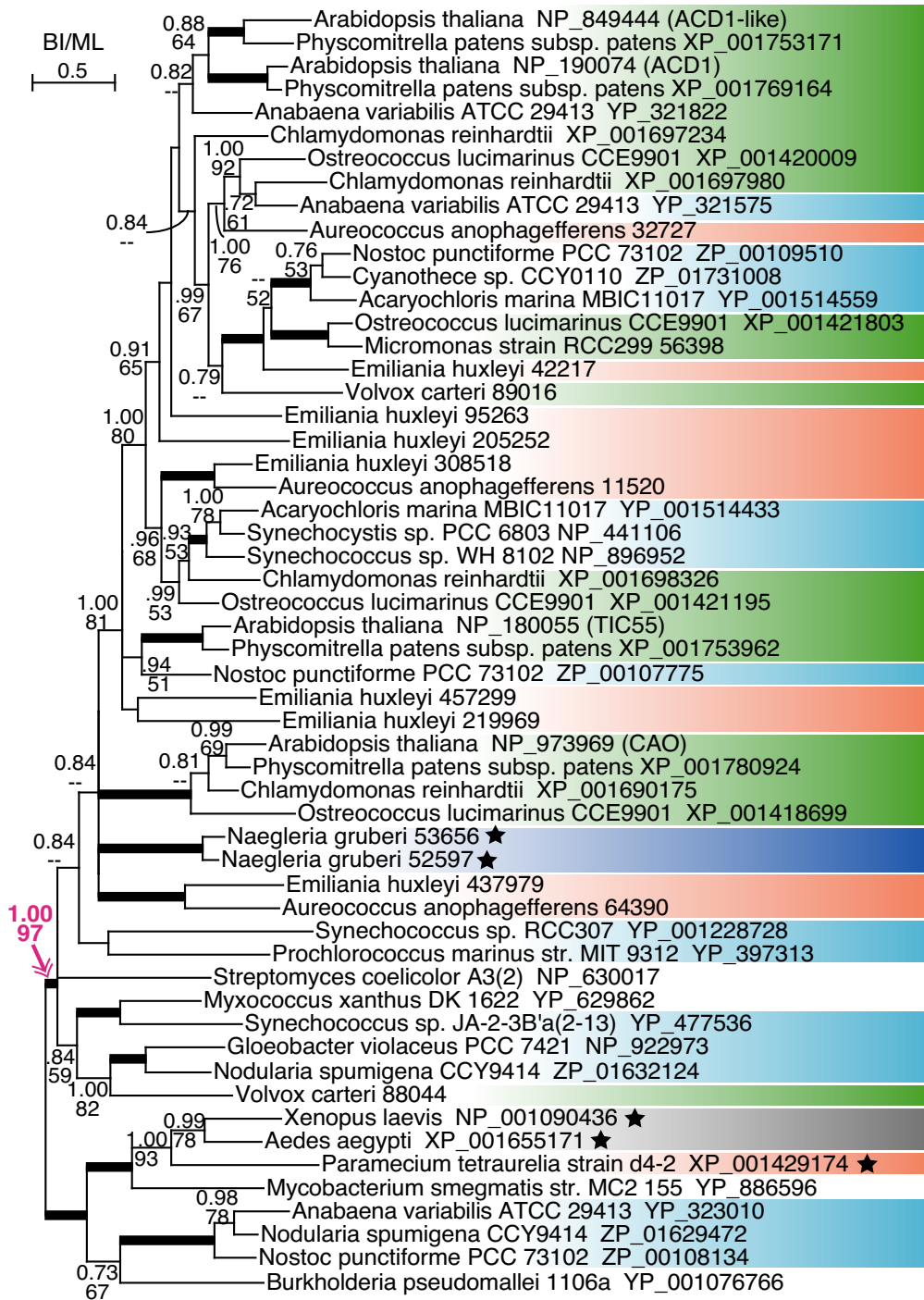
Figure 5
Naegleria **genes are a member of the multiple gene families of TIC55-like oxidoreductase genes**. See legend for figure 2.

Table 1: Summary of eukaryote-type genes with cyanobacterial ancestry identified in this study

Gene family	Pathway	bikonts Chromalveolata *P. ramorum*	Excavata *N. gruberi*	unikonts Amoebozoa *D. discoideum*	Opisthokonta *S. cerevisiae*	Opisthokonta *M. brevicollis*
Uroporphyrin III methyltransferase	porphyrin	51635	-	-	-	XP_001742170
Cobalamin-independent methionine synthase	methionine	72019	-	-	-	-
Amino acid decarboxylase	amino acid	-	36109	-	-	-
TIC55-like oxidoreductase	unknown	-	52597	-	-	-
Folate/biopterin transporter	folate	72218	-	-	-	-
6-phosphogluconate dehydrogenase	pentose phosphate	71783	30694	-	-	-
Cobalamin synthesis protein	cobalamin	85610	38446	-	-	XP_001746731
Oligopeptidase	unknown	54177	-	-	-	-
YCF45	unknown	83996	2396	-	-	-
Glycerate kinase	glyoxylate	94130	-	-	YGR205w	-
Amino acid aminotransferase	amino acid	-	2119	-	-	XP_001749475
Glyoxalase I family protein-like	unknown	-	29304	-	-	XP_001750995

JGI gene ID (*P. ramorum* and *N. gruberi*), gene name (*S. cerevisiae*) and accession numbers (*M. brevicollis*) are listed.

Interestingly, while *Phytophthora ycf21* homologs, probably transferred from a relative of the extant cyanobacterial species via LGT, were placed within the cyanobacterial gene clade, the ciliate *Tetrahymena ycf21* homolog showed affinity to Archaeplastida (Additional file 11). This gene was neither found in the *Paramecium* genome nor in the list of the recently identified algal genes in ciliates [27].

We found that Uroporphyrin III methyltransferase gene homologs (Figure 2) consisted of two large subfamilies of genes with cyanobacterial ancestry, and that oomycete genes were included only in one of them. Given that both subfamilies include green plants, red algae, chromalveolates and cyanobacteria, it is likely that they diverged within the ancestral cyanobacteria and transferred into eukaryotic hosts via primary and secondary endosymbioses. Both of the subfamilies were concurrently present in the cyanobacterial and green algal genomes. In land plants, red algae, diatoms, haptophytes and the plastid-lacking oomycetes, one of the subfamilies might be lost along with the loss of the plastid. The *Thalassiosira* homolog formed a monophyletic group with green plants, rather than red algae, suggesting that it was acquired independently of the secondary plastid of the red lineage. In this study, the bacteriovorous choanoflagellate *Monosiga brevicollis* gene and the proteobacterial genes (*Gluconobacter*, *Alteromonas* and *Nitrosomonas*) were treated as 'apparently LGT-derived genes', incongruously showing affinities to photosynthetic bikonts [33].

Genes encoding cobalamin-independent methionine synthase in green and red algae, diatoms, and oomycetes formed a monophyletic group with cyanobacterial homologs, while the land plants and the red alga *Cyanidioschyzon* homologs were placed in different clades unrelated to cyanobacteria (Figure 3). Close association between diatom and oomycete genes suggested the deep ancestry of the genes in the chromalveolate lineage. We failed to find the homologs in the prasinophytes *Ostreococcus* and *Micromonas*, suggesting that this gene family was dispensable in some plant lineages.

One of the genes with cyanobacterial ancestry found in *N. gruberi* is pyridoxal-dependent amino acid decarboxylase gene (Figure 4). The tree indicated that green plants were split into different eukaryotic clades. *Naegleria* and chromalveolate genes showed robust monophyly with green plants, included in a cyanobacterial gene clade. The tree showed that land plants possessed another subfamily, associated with red algal and fungal genes, apparently of non-cyanobacterial origin. We also identified genes with cyanobacterial ancestry from *Naegleria* in an oxidoreductase gene family that included genes encoding Rieske iron-sulfur cluster 55 kDa protein of chloroplast inner membrane translocon (TIC55), chlorophyll a oxidase (CAO), Lethal-leaf spot 1 (LLS1, which is synonymous with pheophorbide a oxygenase (PAO)) and accelerated cell death 1 (ACD1) (Figure 5) [34,35]. All the members of this family in land plants were hypothesized to be located at the inner membrane of the chloroplast, and to be involved in chlorophyll metabolism [34]. The phylogenetic tree of the *TIC55*-like gene family showed intricate distribution of cyanobacterial, green plant and chromalveolate genes.

In other trees of the genes identified in this study (Additional files 2, 3, 4, 5, 6, 7, 8 and 9), gene clades with cyanobacterial ancestry were mostly composed of bikonts genes, besides the choanoflagellate *M. brevicollis* genes (see Discussion).

Discussion

We identified eight and seven genes with cyanobacterial ancestry in the genome sequences of the oomycete *P. ramorum* and the heterolobosean *N. gruberi*, respectively

(Table 1). It was reported that the apicomplexan *Cryptosporidium* 'recently' lost their secondary plastid, and retained two to seven putative plastid-derived genes in the genome [36]. This number is comparable to our result of the gene mining study using oomycete and heterolobosean genomes. In addition, our system resolved the hidden diversity of the gene family repertoire in eukaryotic genomes by one-by-one gene phylogenies.

Secondary EGT scenario

Although the phylogenetic positions of Cryptophyceae and Haptophyta are still debatable [e.g. [17,37-41]], the chromalveolate hypothesis has been reinstated to support the evolutionary scenario that the plastid-lacking protists oomycetes and ciliates once might have had a plastid [27]. According to this hypothesis, the genes with cyanobacterial ancestry found in the oomycete genomes were acquired via secondary EGT in the common ancestor of Chromalveolata, from the red algal ancestor of secondary plastids. This explanation is also applicable under the alternative hypothesis for chromalveolate plastids, which proposes that a tertiary endosymbiont of the haptophyte/cryptophyceae lineage is the origin of the stramenopile/alveolate plastids [22]. The phylogenetic tree of the photorespiratory glycerate kinase genes, suggesting the red algal origin of the *Phytophthora* genes (Additional file 7), is consistent with the chromalveolate hypothesis. However, several other gene trees in this study showed oomycete genes with green lineage affinity, not red algae (e.g. Additional files 2, 3 &4). Recently, Frommolt et al. [42] demonstrated that, out of 16 genes involved in carotenoid biosynthesis from chromalveolate algae, one third (5/16) of plastid-targeted, nuclear-encoded genes are most closely related to green algal homologs. Reyes-Prieto, Moustafa and Bhattacharya [27] identified 16 genes of possible algal origin in the ciliates *Tetrahymena thermophila* and *Paramecium tetraurelia*, and 7/16 of their trees show a close relationship between green plants and Chromalveolata. Frommolt et al. [42] attributed the close relationships between green plants and chromalveolate genes to the secondary endosymbiosis of an ancient green plant (e.g. prasinophyte), based on the hypothesis on the monophyly of the Archaeplastida [16,40]. This explanation might be also applicable to the plant-like genes in ciliates [27].

While Heterolobosea and Euglenozoa are often united as the morphologically defined taxon, Discicristata, within Excavata [10], recent morphological and molecular phylogenetic analyses suggest that the heteroloboseans (e.g. *Naegleria*) never possessed the secondary plastid of green lineage and share the same origin with Euglenida [43]. Molecular phylogenetic analyses showed that Excavata is separated from other secondary plastid-containing eukaryotes (Chromalveolata and Rhizaria) [18,40]. There-

fore, it is unlikely that the genes with cyanobacterial ancestry found in the heterolobosean nuclear genomes originated from the plastid cognate with any known secondary plastids in extant photosynthetic eukaryotes. The amino acid decarboxylase gene (Figure 4) and the *gnd* gene (Additional file 3) [21] trees demonstrated the presence of genes with cyanobacterial ancestry in other heterolobosean species than *N. gruberi*, suggesting that the ancestor of the genus *Naegleria* possessed this gene family. Furthermore, although ML bootstrap support or Bayesian posterior probability (BI) values were not always sufficient, the *Naegleria* genes occupy relatively basal phylogenetic positions within the bikonts clade in all seven trees (Figures 4 and 5; Additional files 3, 4, 6, 8, and 9). Thus it is possible that the genes with cyanobacterial ancestry were introduced en bloc in the ancestor of Heterolobosea, via a batch gene transfer, in a concerted manner. One possible origin of such a concerted gene transfer is secondary EGT from a photosynthetic eukaryote with a basal phylogenetic position within bikonts. However, as discussed above, it is unlikely that Heterolobosea experienced secondary endosymbiosis and acquired genes common to the extant secondary plastid-containing eukaryotes via secondary EGT.

Ancient eukaryote-to-eukaryote LGT or primary EGT scenarios

Alternatively, we can argue for two other explanations: a concerted eukaryote-to-eukaryote LGT scenario or a more ancient primary EGT scenario. The *Naegleria* genes with cyanobacterial ancestry shown in Table 1 are basally positioned within bikonts, but not intruding into any of gene clades from extant photosynthetic eukaryotes (Figures 4 and 5; Additional files 3, 4, 6, 8, and 9). Thus, if we assume that these genes were acquired via non-endosymbiotic LGT, they may originate from unknown ancient photosynthetic lineages basally positioned within bikonts. Meanwhile, under the primary EGT scenario, in which the primary endosymbiosis occurred in the common ancestor of bikonts (Figure 1) [[19-21], but see Ref. [9] for further discussion on the root of eukaryotic tree of life], ancient primary EGT occurred much earlier than the conventional hypothesis, from the cyanobacterium-like prokaryote to the common ancestor of bikonts. Primary plastids were subsequently lost in many lineages of bikonts, except for the Archaeplastida lineages, but some genes originating from the cyanobacterial ancestor of the primary plastids have been retained in the nuclear genomes of the plastid-lacking lineages of bikonts (Figures 1 and 6). The loss of the plastid might have triggered the loss of genes that specifically functioned within the plastid. Only a portion of the plastid-derived genes, which we can find now in the plastid-lacking protist genomes, might have escaped from or survived through eliminative pressure in a lineage-specific manner, by acquiring addi-
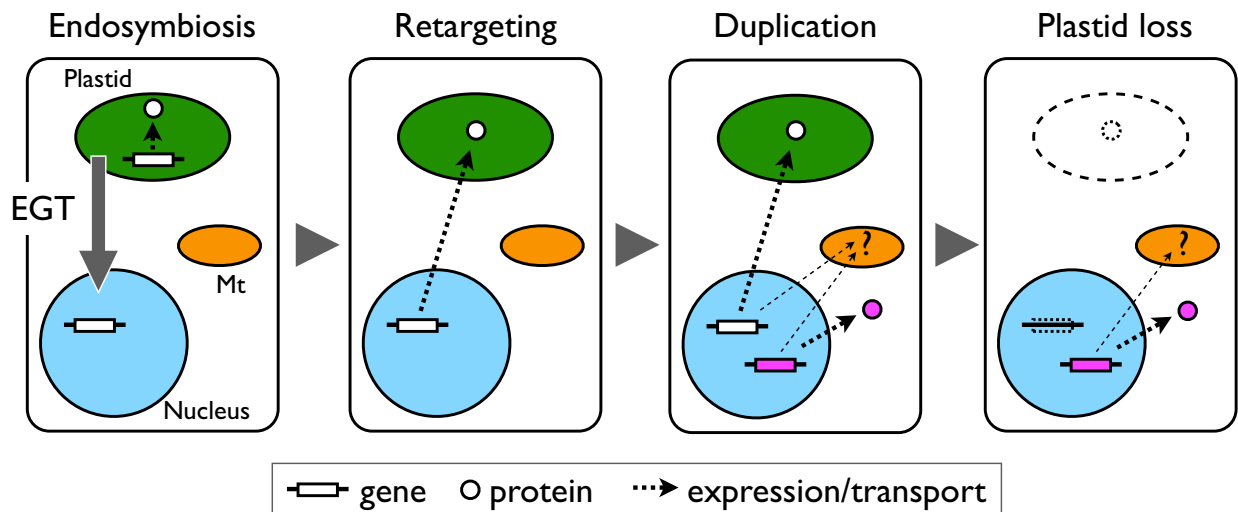
Figure 6
An evolutionary history of the genes with cyanobacterial ancestry. Thick continuous arrows represent gene flow via EGT. Thin broken arrows indicate gene expression or intracellular transport into organelles. Dashed line circles and boxes indicate that they have been lost in the evolutionary history. Note that the genes with cyanobacterial ancestry (white), which had been derived from the plastid genome via EGT, were retargeted into the plastid. After rounds of gene family duplication, some genes (magenta) gained additional functions in other cellular compartments (cytosol, mitochondrion, etc.). In some plastid-lacking protists, a number of genes were retained in the nuclear genomes after the plastid loss events. Mt, mitochondrion.

tional functions with other components and/or in other cellular compartments. This might account for the observed punctate distribution of gene families among the eukaryotes [44,45].

Recently, a hypothesis for the non-monophyly of Archaeplastida was proposed based on the phylogenetic analyses of slowly evolving nuclear-encoded genes [17,19]. This non-monophyly hypothesis could be also considered within the scope of the primary EGT scenario. It is notable that a number of the trees in this study (Figure 2; Additional files 2, 3, 4, 6, and 8) showed intriguing topologies, depicting the split of Archaeplastida and inclusion of Chromalveolata and Excavata genes within it, as shown in the previously reported multiple slowly-evolving gene phylogeny [19] and *gnd* gene phylogeny [20,21]. These results are consistent with the hypothesis for the non-monophyly of Archaeplastida, and suggest that the oomycete and heterolobosean genes with cyanobacterial ancestry might reflect the host nuclear genome phylogeny. On the other hand, the genes found in the marine choanoflagellate *M. brevicollis* were positioned within the bikonts clade, but not associated with the genes from other Opisthokonta relatives (Metazoa and fungi), suggesting that the tree topologies were probably not reflective of the host phylogeny [46] but eukaryote-to-eukaryote LGT (Figure 2; Additional files 4, 8, and 9). No gene with cyanobacterial ancestry was found in *D. discoi-*

deum (Amoebozoa), and only one gene in *S. cerevisiae* (Opisthokonta). These results are also consistent with the ancient primary EGT scenario.

A photorespiratory gene with cyanobacterial ancestry in fungi
Our analysis using the genome data of the budding yeast *S. cerevisiae* identified one gene with cyanobacterial ancestry, encoding the glycerate kinase for photorespiration (Additional file 7). Given that photorespiration is essential for cyanobacteria and plants, it is likely that the glycerate kinases in plants and cyanobacteria are phylogenetically and physiologically related to photorespiration [47,48]. A previous study on glycerate kinases showed that, regardless of the complete absence of photorespiratory metabolism in fungi, the gene product from the budding yeast *Saccharomyces* showed similar enzymatic activity and substrate specificity compared with the *Arabidopsis* gene, suggesting that the plant and fungal genes catalyze the same reaction in different contexts of the metabolic pathway [47]. Another example of plant-type genes in fungi was reported in a phylogenetic study of the genes encoding high-affinity nitrate transporter NRT2, which suggested that fungi probably acquired the NRT2 genes via LGT from one of the chromalveolate lineages [49]. Meanwhile, our data showed that the fungal clade was located outside the clade of plants plus oomycetes (Additional file 7), suggesting that fungal glycerate

kinase genes with cyanobacterial ancestry likely originated from an LGT event from an ancestor of cyanobacteria, or eukaryote-to-eukaryote LGT from an ancestor of Archaeplastida (or bikonts). One likely explanation for the presence of photorespiratory genes in oomycetes is that the ancestor of Chromalveolata possessed this gene family, but some photosynthetic descendants lost this gene family or replaced it with other genes during the course of lineage-dependent customization of photorespiratory pathways [[50,51]; for discussion on carbon assimilation in diatoms], while oomycetes retained the genes without any replacement.

Gene family expansion and differential reduction

Another conclusion of this analysis is that rounds of gene family expansion and selective reduction are important factors in making eukaryotic genome phylogeny look like a complicated mosaic (Figure 6). It is likely that the alteration of gene family repertoire contributed to the restructuring of the intracellular metabolome and a reduction of the dispensable gene families. Our data showed that all the genes identified in this study were members of multiple gene families. Algae and plastid-lacking protists retained only members of subfamilies (e.g. Figure 2 and Additional file 8), suggesting that the punctate distribution might be a corollary of the common mechanism by which genes with cyanobacterial ancestry were retained in their genomes. The presence of genes from multiple subfamilies in one organism supports this idea (e.g. two Uroporphyrin III methyltransferase subfamilies in prasinophytes and *Volvox* in Figure 2). Discontinuous loss or gain of a metabolic pathway in a lineage might be another factor in punctate distribution; e.g. the oxidative pentose pathway, and the cyanobacterial *gnd* genes functioning therein, were present in most bikonts but lost in the ciliate *Tetrahymena* [21,52]. A recent study on pyridoxal-dependent amino acid aminotransferase reported that, besides the ancestrally eukaryotic enzymes, land plants possess a distinct subfamily of prokaryote-type chloroplast-targeted enzymes [53]. Our data with richer taxon sampling identified another prokaryote-type subfamily with cyanobacterial ancestry (Additional file 8), illustrating the hidden evolutionary diversity of protist and algal metabolomes.

Future prospects

Our results showed that many genes with cyanobacterial ancestry identified in this study were found only in complete genome sequences, suggesting that these genes might be difficult to discover by expressed sequence tag (EST) library sequencing, probably due to the low-level expression of these genes. Although the whole genome data from excavate parasites (e.g. *Trypanosoma*, *Giardia* and *Trichomonas*) are available, they seem to be unsuited for the gene mining study because of the unusual nucle-

otide substitutions (see Methods). At the stage of starting the present gene mining study, *N. gruberi* was the only species with whole genome data released within the non-parasitic excavates, and thereby the excavate genes with cyanobacterial ancestry were mostly from *N. gruberi*. More genome data from plastid-lacking protists from Excavata and Rhizaria as well as Archaeplastida, especially red algae and glaucophytes, are needed to unravel the evolutionary history of plastids, and plastid-lacking protists.

Conclusion

The comparative analyses of the genome sequence data of the plastid-lacking eukaryotes demonstrated the potentially significant contributions of ancestral or extant cyanobacteria to the eukaryotic genomes, which probably occurred via LGT or ancient primary EGT events. Furthermore, the automated phylogenetic analyses revealed the diversity and punctate distribution of gene families within the genomes in the unicellular microbes. More genome data of the plastid-lacking Excavata and Rhizaria will make the evolutionary history clear and support our hypotheses.

Methods

Data preparation

The genome sequence data of *P. ramorum*, *N. gruberi* and *M. brevicollis* was produced by the US Department of Energy Joint Genome Institute (JGI) [54]. *D. discoideum* genome data (9 Nov 2007) at dictyBase [55] and *S. cerevisiae* genome data [56] were used for phylogenetic analysis. Red algal data were retrieved from the *Cyanidioschyzon merolae* [57], *Galdieria sulphuraria* [58] genome databases, and other algal data were from *Aureococcus anophagefferens*, *Emiliania huxleyi*, *Micromonas pusilla*, *Micromonas* sp. RCC299, *Ostreococcus tauri*, *Ostreococcus* sp. RCC809, *Phaeodactylum tricornutum*, *Phytophthora sojae*, *Thalassiosira pseudonana* and *Volvox carteri* genome databases on JGI. EST sequences of several protists were obtained from TBestDB [59] and all other sequences were from the NCBI GenBank refseq database [60]. We excluded amitochondrial and/or parasitic eukaryotes, which might cause long branch attraction due to unusual nucleotide substitutions [61,62]. Fragments of *N. fowleri* amino acid decarboxylase gene [DDBJ: AB491948] were amplified from genomic DNA using degenerated primers based on the conserved amino acid motif YHHFGYP for the forward primer (TAYCAYCAYTTIGGITAYCC) and WQLACEG for the reverse primer (CCYTCRCAIGCIARYTGCCA). PCR products were directly sequenced using an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) with a BigDye Terminator Cycle Sequencing Ready Reaction kit v. 3.1 (Applied Biosystems).

Phylogenomic analysis

A genome-wide phylogenetic program was made with several bench-made BioRuby scripts (Additional file 1), referring to the previously reported phylogenomic pipeline used in the macronuclear genome analysis of *Tetrahymena thermophila* [63]. For the first screening, query amino acid sequences were automatically subjected to BLAST searching using NCBI netblast [64] and EFetch utilities [65], extracting the genes showing the highest E-value to a cyanobacterial counterpart among 'Bacteria' by BLASTP. For the second step, these genes were subjected to BLASTP analysis against 'refseq-protein' to fetch homologous sequences with E-values less than 0.001, up to 500 hits at a maximum. Multiple alignments were then performed using MUSCLE [66], which automatically removed ambiguously aligned sites or sequences with too many gaps. Bootstrapped neighbor-joining trees were produced using QuickTree [67]. Trees were output in the PostScript format using the newicktops program in the NJplot package [68] with sizes and colors of OTU names modified according to the NCBI taxonomy database [69] to simplify the subsequent visual checking process. Genomes of several bacterial genera were intensively sequenced and many homologous sequences from closely related species and strains (e.g. *Escherichia*, *Bacillus*) appeared on the trees. To diminish the sampling bias, the output files of QuickTree were also used to parse tree topology and detect a monophyletic clade exclusively composed of OTUs from a single genus using Bio::Tree class methods in BioRuby scripts. One representative OTU was automatically selected in such single-genus clades, the other OTUs were removed, and the trees were re-constructed for visual checking. In addition to the automatic process, trees for genes listed in the putative photosynthetic endosymbiont-derived genes [28], but not detected in our analysis, were manually re-constructed. Non-cyanobacterial prokaryotic genes taxonomically unrelated to, but placed within, the cyanobacterial clade were interpreted as 'apparently LGT-derived genes' with cyanobacterial ancestry.

Candidate cyanobacteria-related genes were manually selected, their homologs were collected from major groups of the three domains of life, and then subjected to multiple protein sequence alignments using MUSCLE. Phylogenetic analyses were performed with a maximum likelihood (ML) method using RAxML [70] and with a Bayesian interference (BI) method using MrBayes [71]. ML and BI were based on the WAG substitution matrix with options of four gamma-distributed rate categories and estimate of invariable sites (plus empirical base frequencies in ML). ML branch support was evaluated with 1000 bootstrap replicates, and BI posterior probability values were calculated from the MCMC run data, which summarized when the average standard deviation of split

frequencies reached less than 0.01. Except for cyanobacterial genes of which no homologs were found in other prokaryotes (e.g. Additional file 2), or of which monophyly was confirmed by previous studies (e.g. Additional file 3), threshold values to assess the monophyly of cyanobacterial gene clades were 50% on ML bootstrap or 0.9 on BI posterior probability values.

Abbreviations

BI: Bayesian posterior probability; EGT: endosymbiotic gene transfer; EST: expressed sequence tag; LGT: lateral gene transfer; ML: maximum likelihood; NJ: neighbor-joining; OUT: Operational Taxonomic Unit; TIC55: Rieske iron-sulfur cluster 55 kDa protein of chloroplast inner membrane translocon.

Authors' contributions

SM and HN conceived the study. SM, MM and KM prepared and analyzed the data. SM and HN drafted the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1
Supplemental Figure 7. *Flow chart of procedures used in the phylogenetic analyses.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-197-S1.pdf]

Additional file 2
Supplemental Figure 8. *MrBayes consensus tree of folate/biopterin transporter genes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-197-S2.pdf]

Additional file 3
Supplemental Figure 9. *MrBayes consensus tree of 6-phosphogluconate dehydrogenase genes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-197-S3.pdf]

Additional file 4
Supplemental Figure 10. *MrBayes consensus tree of cobalamin synthesis protein genes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-197-S4.pdf]

Additional file 5
Supplemental Figure 11. *MrBayes consensus tree of oligopeptidase genes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-197-S5.pdf]

Additional file 6

Supplemental Figure 12. *MrBayes consensus tree of YCF45 genes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-
2148-9-197-S6.pdf]

Additional file 7

Supplemental Figure 13. *MrBayes consensus tree of glycerate kinase
genes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-
2148-9-197-S7.pdf]

Additional file 8

Supplemental Figure 14. *MrBayes consensus tree of amino acid ami-
notransferase genes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-
2148-9-197-S8.pdf]

Additional file 9

Supplemental Figure 15. *ML consensus tree of glyoxalase I family pro-
tein-like genes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-
2148-9-197-S9.pdf]

Additional file 10

Supplemental Figure 16. *MrBayes consensus tree of phosphoadenosine
phosphosulfate reductase genes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-
2148-9-197-S10.pdf]

Additional file 11

Supplemental Figure 17. *MrBayes consensus tree of YCF21 genes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-
2148-9-197-S11.pdf]

Additional file 12

Supplemental Figure 18. *MrBayes consensus tree of hypothetical protein
genes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-
2148-9-197-S12.pdf]

References

1. Timmis JN, Ayliffe MA, Huang CY, Martin W: **Endosymbiotic gene
   transfer: organelle genomes forge eukaryotic chromosomes.**
   Nat Rev Genet 2004, **5**:123-135.
2. Delwiche CF: **Tracing the thread of plastid diversity through
   the tapestry of life.** *Am Nat* 1999, **154**:164-177.
3. Bhattacharya D, Yoon HS, Hackett JD: **Photosynthetic eukaryo-
   tes unite: endosymbiosis connects the dots.** *Bioessays* 2004,
   26:50-60.
4. Baldauf SL: **The deep roots of eukaryotes.** *Science* 2003,
   300:1703-1706.
5. Stechmann A, Cavalier-Smith T: **Rooting the eukaryote tree by
   using a derived gene fusion.** *Science* 2002, **297**:89-91.
6. Stechmann A, Cavalier-Smith T: **The root of the eukaryote tree
   pinpointed.** *Curr Biol* 2003, **13**:665-666.
7. Nozaki H, Matsuzaki M, Misumi O, Kuroiwa H, Higashiyama T,
   Kuroiwa T: **Phylogenetic implications of the CAD complex
   from the primitive red alga *Cyanidioschyzon merolae* (Cya-
   nidiales, Rhodophyta).** *J Phycol* 2005, **41**:652-657.
8. Richards TA, Cavalier-Smith T: **Myosin domain evolution and the
   primary divergence of eukaryotes.** *Nature* 2005,
   436:1113-1118.
9. Roger AJ, Simpson AGB: **Revisiting the root of the eukaryote
   tree.** *Curr Biol* 2009, **19**:165-167.
10. Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta
    JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY,
    Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH,
    Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge
    SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor MF: **The
    new higher level classification of eukaryotes with emphasis
    on the taxonomy of protists.** *J Eukaryot Microbiol* 2005,
    52:399-451.
11. Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyag-
    ishima SY, Mori T, Nishida K, Yagisawa F, Nishida K, Yoshida Y,
    Nishimura Y, Nakao S, Kobayashi T, Momoyama Y, Higashiyama T,
    Minoda A, Sano M, Nomoto H, Oishi K, Hayashi H, Ohta F, Nishizaka
    S, Haga S, Miura S, Morishita T, Kabeya Y, Terasawa K, Suzuki Y, Ishii
    Y, *et al.*: **Genome sequence of the ultrasmall unicellular red
    alga *Cyanidioschyzon merolae* 10D.** *Nature* 2004, **428**:653-657.
12. Reyes-Prieto A, Bhattacharya D: **Phylogeny of calvin cycle
    enzymes supports plantae monophyly.** *Mol Phylogenet Evol*
    2007, **45**:384-391.
13. Tyra HM, Linka M, Weber AP, Bhattacharya D: **Host origin of plas-
    tid solute transporters in the first photosynthetic eukaryo-
    tes.** *Genome Biol* 2007, **8**:R212.
14. Larkum AW, Lockhart PJ, Howe CJ: **Shopping for plastids.** *Trends
    Plant Sci* 2007, **12**:189-195.
15. Stiller JW: **Plastid endosymbiosis, genome evolution and the
    origin of green plants.** *Trends Plant Sci* 2007, **12**:391-396.
16. Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G,
    Löffelhardt W, Bohnert HJ, Philippe H, Lang BF: **Monophyly of pri-
    mary photosynthetic eukaryotes: green plants, red algae,
    and glaucophytes.** *Curr Biol* 2005, **15**:1325-1330.
17. Nozaki H, Iseki M, Hasegawa M, Misawa K, Nakada T, Sasaki N,
    Watanabe M: **Phylogeny of primary photosynthetic eukaryo-
    tes as deduced from slowly evolving nuclear genes.** *Mol Biol
    Evol* 2007, **24**:1592-1595.
18. Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger
    AJ: **Phylogenomic analyses support the monophyly of Exca-
    vata and resolve relationships among eukaryotic "super-
    groups".** *Proc Nat Acad Sci USA* 2009, **106**:3859-3864.
19. Nozaki H: **A new scenario of plastid evolution: plastid primary
    endosymbiosis before the divergence of the "Plantae,"
    emended.** *J Plant Res* 2005, **118**:247-255.
20. Andersson JO, Roger AJ: **A cyanobacterial gene in nonphoto-
    synthetic protists – an early chloroplast acquisition in
    eukaryotes?** *Curr Biol* 2002, **12**:115-119.
21. Maruyama S, Misawa K, Iseki M, Watanabe M, Nozaki H: **Origins of
    a cyanobacterial 6-phosphogluconate dehydrogenase in plas-
    tid-lacking eukaryotes.** *BMC Evol Biol* 2008, **8**:151.
22. Sanchez-Puerta MV, Delwiche CF: **A hypothesis for plastid evolu-
    tion in chromalveolates.** *J Phycol* 2008, **44**:1097-1107.
23. Cavalier-Smith T: **Principles of protein and lipid targeting in
    secondary symbiogenesis: euglenoid, dinoflagellate, and spo-
    rozoan plastid origins and the eukaryote family tree.** *J
    Eukaryot Microbiol* 1999, **46**:347-366.

24. Foth BJ, McFadden GI: **The apicoplast: a plastid in *Plasmodium falciparum* and other apicomplexan parasites.** *Int Rev Cytol* 2003, **224**:57-110.
25. Matsuzaki M, Kuroiwa H, Kuroiwa T, Kita K, Nozaki H: **A cryptic algal group unveiled: a plastid biosynthesis pathway in the oyster parasite *Perkinsus marinus*.** *Mol Biol Evol* 2008, **25**:1167-1179.
26. Slamovits CH, Keeling PJ: **Plastid-derived genes in the non-photosynthetic alveolate *Oxyrrhis marina*.** *Mol Biol Evol* 2008, **25**:1297-1306.
27. Reyes-Prieto A, Moustafa A, Bhattacharya D: **Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic.** *Curr Biol* 2008, **18**:956-962.
28. Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RH, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, Chapman J, Damasceno CM, Dorrance AE, Dou D, Dickerman AW, Dubchak IL, Garbelotto M, Gijzen M, Gordon SG, Govers F, Grunwald NJ, Huang W, Ivors KL, Jones RW, Kamoun S, Krampis K, Lamour KH, Lee MK, McDonald WH, Medina M, *et al.*: **Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis.** *Science* 2006, **313**:1261-1266.
29. Koski LB, Golding GB: **The closest BLAST hit is often not the nearest neighbor.** *J Mol Evol* 2001, **52**:540-542.
30. Nozaki H, Matsuzaki M, Misumi O, Kuroiwa H, Hasegawa M, Higashiyama T, Shin-I T, Kohara Y, Ogasawara N, Kuroiwa T: **Cyanobacterial genes transmitted to the nucleus before divergence of red algae in the Chromista.** *J Mol Evol* 2004, **59**:103-113.
31. Schuster FL, Visvesvara GS: **Free-living amoebae as opportunistic and non-opportunistic pathogens of humans and animals.** *Int J Parasitol* 2004, **34**:1001-1027.
32. Simpson AG, Perley TA, Lara E: **Lateral transfer of the gene for a widely used marker, alpha-tubulin, indicated by a multiprotein study of the phylogenetic position of Andalucia (Excavata).** *Mol Phylogenet Evol* 2008, **47**:366-377.
33. Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ: **Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*.** *Proc Natl Acad Sci USA* 2003, **100**:7678-7683.
34. Gray J, Wardzala E, Yang M, Reinbothe S, Haller S, Pauli F: **A small family of LLS1-related non-heme oxygenases in plants with an origin amongst oxygenic photosynthesizers.** *Plant Mol Biol* 2004, **54**:39-54.
35. Gross J, Bhattacharya D: **Revaluating the evolution of the Toc and Tic protein translocons.** *Trends Plant Sci* 2009, **14**:13-20.
36. Huang J, Mullapudi N, Lancto CA, Scott M, Abrahamsen MS, Kissinger JC: **Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*.** *Genome Biol* 2004, **5**:R88.
37. Moreira D, Heyden S von der, Bass D, López-García P, Chao E, Cavalier-Smith T: **Global eukaryote phylogeny: combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata.** *Mol Phylogenet Evol* 2007, **44**:255-266.
38. Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rümmele SE, Bhattacharya D: **Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates.** *Mol Biol Evol* 2007, **24**:1702-1713.
39. Patron NJ, Inagaki Y, Keeling PJ: **Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages.** *Curr Biol* 2007, **17**:887-891.
40. Burki F, Shalchian-Tabrizi K, Pawlowski J: **Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes.** *Biol Lett* 2008, **4**:366-369.
41. Minge M, Silberman JD, Orr R, Cavalier-Smith T, Shalchian-Tabrizi K, Burki F, Skjæveland Å, Jakobsen KS: **Evolutionary position of breviate amoebae and the primary eukaryote divergence.** *Proc Biol Sci* 2009, **276**:597-604.
42. Frommolt R, Werner S, Paulsen H, Goss R, Wilhelm C, Zauner S, Maier UG, Grossman AR, Bhattacharya D, Lohr M: **Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis.** *Mol Biol Evol* 2008, **25**:2653-2667.
43. Leander BS: **Did trypanosomatid parasites have photosynthetic ancestors?** *Trends Microbiol* 2004, **12**:251-258.
44. Keeling PJ, Inagaki Y: **A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1alpha.** *Proc Natl Acad Sci USA* 2004, **101**:15380-15385.
45. Rogers MB, Watkins RF, Harper JT, Durnford DG, Gray MW, Keeling PJ: **A complex and punctate distribution of three eukaryotic genes derived by lateral gene transfer.** *BMC Evol Biol* 2007, **7**:89.
46. Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, Cavalier-Smith T: **Multigene phylogeny of choanozoa and the origin of animals.** *PLoS ONE* 2008, **3**:e2098.
47. Boldt R, Edner C, Kolukisaoglu U, Hagemann M, Weckwerth W, Wienkoop S, Morgenthal K, Bauwe H: **D-GLYCERATE 3-KINASE, the last unknown enzyme in the photorespiratory cycle in Arabidopsis, belongs to a novel kinase family.** *Plant Cell* 2005, **17**:2413-2420.
48. Eisenhut M, Ruth W, Haimovich M, Bauwe H, Kaplan A, Hagemann M: **The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants.** *Proc Natl Acad Sci USA* 2008, **105**:17199-17204.
49. Slot JC, Hallstrom KN, Matheny PB, Hibbett DS: **Diversification of NRT2 and the origin of its fungal homolog.** *Mol Biol Evol* 2007, **24**:1731-43.
50. Wilhelm C, Büchel C, Fisahn J, Goss R, Jakob T, Laroche J, Lavaud J, Lohr M, Riebesell U, Stehfest K, Valentin K, Kroth PG: **The regulation of carbon and nutrient assimilation in diatoms is significantly different from green algae.** *Protist* 2006, **157**:91-124.
51. Roberts K, Granum E, Leegood RC, Raven JA: **C3 and C4 pathways of photosynthetic carbon assimilation in marine diatoms are under genetic, not environmental, control.** *Plant Physiol* 2007, **145**:230-235.
52. Eldan MB, Blum J: **Presence of Nonoxidative Enzymes of the Pentose Phosphate Shunt in Tetrahymena.** *J Eurkaryot Microbiol* 1975, **22**:145-149.
53. de la Torre F, De Santis L, Suárez MF, Crespillo R, Cánovas FM: **Identification and functional analysis of a prokaryotic-type aspartate aminotransferase: implications for plant amino acid metabolism.** *Plant J* 2006, **46**:414-425.
54. **DOE Joint Genome Institute** [http://www.jgi.doe.gov/]
55. **DictyBase: An Online Informatics Resource for *Dictyostelium*** [http://dictybase.org/]
56. **MIPS Comprehensive Yeast Genome Database** [http://mips.gsf.de/genre/proj/yeast/]
57. ***Cyanidioschyzon merolae* Genome Database** [http://merolae.biol.s.u-tokyo.ac.jp/]
58. ***Galdieria sulphuraria* Genome Database** [http://genomics.msu.edu/galdieria/]
59. **Taxonomically broad EST database TBestDB** [http://tbestdb.bcm.umontreal.ca/]
60. **National Center for Biotechnology Information (NCBI)** [http://www.ncbi.nlm.nih.gov/]
61. Stiller JW, Duffield EC, Hall BD: **Amitochondriate amoebae and the evolution of DNA-dependent RNA polymerase II.** *Proc Natl Acad Sci USA* 1998, **95**:11769-11774.
62. Stiller JW, Riley J, Hall BD: **Are red algae plants? A critical evaluation of three key molecular data sets.** *J Mol Evol* 2001, **52**:527-539.
63. Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM, Tallon LJ, Delcher AL, Salzberg SL, Silva JC, Haas BJ, Majoros WH, Farzad M, Carlton JM, Smith RK, Garg J, Pearlman RE, Karrer KM, Sun L, Manning G, Elde NC, Turkewitz AP, Asai DJ, Wilkes DE, Wang Y, Cai H, *et al.*: **Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote.** *PLoS Biol* 2006, **4**:e286.
64. **NCBI netblast** [http://www.ncbi.nlm.nih.gov/BLAST/download.shtml]
65. **NCBI EFetch utilities** [http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?]
66. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792-1797.
67. Howe K, Bateman A, Durbin R: **QuickTree: building huge Neighbour-Joining trees of protein sequences.** *Bioinformatics* 2002, **18**:1546-1547.
68. Perrière G, Gouy M: **WWW-query: an on-line retrieval system for biological sequence banks.** *Biochimie* 1996, **78**:364-369.
69. **NCBI taxonomy database** [http://www.ncbi.nlm.nih.gov/Taxonomy]

70. Stamatakis A, Hoover P, Rougemont J: **A rapid bootstrap algorithm for the RAxML Web servers.** *Syst Biol* 2008, **57:**758-771.
71. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17:**754-755.