**Complex evolution of two types of cardiolipin synthase in the eukaryotic lineage
stramenopiles**

Fumiya Noguchi^{a,b}, Goro Tanifuji^c, Matthew W. Brown^{d,e}, Katsunori Fujikura^b, Kiyotaka
Takishita^{a,b,*}

^aGraduate School of Marine Science and Technology, Tokyo University of Marine
Science and Technology, Minato-ku, Tokyo, Japan

^bJapan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka,
Kanagawa, Japan

^cGraduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba,
Ibaraki, Japan

^dDepartment of Biological Sciences, Mississippi State University, Mississippi, USA

^eInstitute for Genomics, Biocomputing & Biotechnology, Mississippi State University,
Mississippi, USA

*Corresponding author: Kiyotaka Takishita, Japan Agency for Marine-Earth Science and
Technology, 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan

Telephone number: +81-46-867-9538

FAX number: +81-46-867-9525

E-mail: takishitak@jamstec.go.jp

Abstract

The phospholipid cardiolipin is indispensable for eukaryotes to activate mitochondria, and it was previously reported that two phylogenetically distinct types of enzyme synthesizing cardiolipin, one with two phospholipase D domains (CLS_pld) and the other with a CDP-alcohol phosphatidyltransferase domain (CLS_cap), are patchily and complementarily distributed at higher taxonomic (e.g., supergroup) levels of eukaryotes. Stramenopiles, one of the major eukaryotic clades, were considered to exclusively possess CLS_cap. However, through our present surveys with genome or transcriptome data from a broad range of stramenopile taxa, species with both CLS_cap and CLS_pld and species with only CLS_pld or CLS_cap were discovered among this group. Because these homologues of CLS_cap and CLS_pld retrieved from stramenopiles were likely inherited from the last eukaryotic common ancestor, it is reasonable to assume that a common ancestor of all stramenopiles harbored both CLS_cap and CLS_pld. Furthermore, based on the robust organismal phylogeny of stramenopiles unveiled with large-scale phylogenetic analyses, the earliest diverging lineage of stramenopiles (including bicosoecids, placidids, etc.) was found to comprise species with both CLS_cap and CLS_pld along with species with only either CLS_cap or CLS_pld. These findings suggest that a common ancestor of the most basal stramenopile lineage retained these two vertically inherited enzymes and that differential losses of either CLS_cap or CLS_pld occurred in this lineage. On the other hand, in the other stramenopile lineage composed of Ochrophyta, Pseudofungi, and Labyrinthulomycetes (to the exclusion of the most basal lineage), only CLS_cap was found, and therefore a common ancestor of these three

groups likely lost CLS_pld. Based on our findings, the evolution of CLS_cap/CLS_pld in stramenopiles appears to be more complex than previously thought.

Keywords

Cardiolipin synthase, Differential gene loss, Eukaryotes, Phylogenomics, Stramenopiles

1. Introduction

Cardiolipin (CL) is an anionic dimeric phospholipid with two phosphate residues and four types of fatty acyl chain. It is widely held that CL in mitochondria plays a pivotal role in stabilizing and/or regulating various mitochondrial proteins such as the respiratory chain complexes/supercomplexes (e.g., Fry and Green, 1981; Eble et al., 1990; Robinson et al., 1990). Furthermore, mitochondrial CL is involved in organizing structure characteristic of mitochondria such as cristae, as well as in aging and apoptosis (e.g., Schug and Gottlieb, 2009; Sakamoto et al., 2012). It is also known that bacterial CL is essential to activate proteins associated with energy metabolism and to properly localize membrane proteins (e.g., Yankovskaya et al., 2003; Gold et al., 2010; Arias-Cartin et al., 2011). Based on these findings, it was originally proposed that CL is exclusively located in mitochondrial (inner) and bacterial plasma membranes (White and Frerman, 1967; Zinser et al., 1991), but recently this type of phospholipid has also been found in the plasma membranes of a very limited number of archaeal species and in the peroxisomal membranes of eukaryotes (Lattanzio et al., 2002; Koga and Morii, 2007; Wriessnegger et al., 2007; Lobasso et al., 2008).

CL is known to be biosynthesized by either of two phylogenetically distinct enzymes: CL synthase (CLS) with two phospholipase D domains, CLS_pld, which synthesizes CL from two molecules of phosphatidylglycerols (PGs) (e.g., Nishijima et al., 1988) or CLS with one CDP-alcohol phosphatidyltransferase domain, CLS_cap, which produces this lipid using a PG and a cytidine diphosphate diacylglycerol (CDP-DAG) as substrates (e.g., Chang et al., 1998; Tuller et al., 1998). It has been believed that CLS_pld and CLS_cap function in bacteria and eukaryotes (mitochondria), respectively (e.g., Schlame, 2008). In contrast to the bacterial-type CL, mitochondrial “immature” CL synthesized by CLS is further remodeled (reacylated), resulting in mature CL generally possessing the same fatty acids at sn-1, 2 sites in one molecule. This eukaryotic CL maturation pathway consists of two steps: in the first step, immature CL is deacylated into monolysocardiolipin (MLCL) with either CL-specific phospholipase (CLD) or calcium-independent phospholipase A2 (iPLA2) beta/gamma (Beranek et al., 2009; Malhotra et al., 2009; Zachman et al., 2010). After the deacylation in the first step, MLCL is then reacylated to generate mature CL with either CoA-independent tafazzin (TAZ) or acylCoA:lysocardiolipin acyltransferase 1 (ALCAT1) (Cao et al., 2004; Gu et al., 2004).

Recently, exceptions to the above-mentioned “simple” hypothesis regarding CLS phylogenetic distribution, in which CLS_pld and CLS_cap are exclusively found in bacteria and eukaryotes, respectively, have been found: actinobacteria and proteobacteria were found to contain CLS_cap-like proteins (Sandoval-Calderón et al., 2009; Tian et al., 2012). In addition, Tian et al. (2012) reported that the eukaryotic supergroups Amoebozoa, Excavata, and Alveolata, a subgroup of the supergroup “SAR” (Adl et al.,

2012), have only CLS_pld (without phylogenetic affiliation to any particular bacterial homologues), while the supergroups Opisthokonta (including animals and fungi) and Archaeplastida (including land plants) along with another SAR subgroup stramenopiles possess only CLS_cap (closely related to alpha-proteobacterial homologues). It should be noted that Tian et al. (2012) analyzed a broad range of eukaryotes including non-model organisms other than animals, yeasts, and plants, resulting in the detection of CLS_pld in eukaryotes. It is also remarkable that both CLS_cap and CLS_pld are patchily and complementarily distributed at higher taxonomic levels of eukaryotes. Based on these findings, Tian et al. (2012) proposed the following evolutionary scenarios: a primitive eukaryotic cell prior to acquisition of mitochondria inherited CLS_pld from its ancient bacterial ancestor, and then mitochondria originating from an endosymbiotic event of an alpha-proteobacterium took CLS_cap to the last eukaryotic common ancestor (LECA), meaning that the LECA harbored both CLS_cap and CLS_pld. Subsequently, either CLS_cap or CLS_pld was differentially lost in various eukaryotic lineages.

We previously analyzed the metabolic characteristics of mitochondrion-related organelles (degenerated mitochondria) in the free-living anaerobic stramenopile *Cantina marsupialis* based on RNA-seq data (Noguchi et al., 2015). Through those analyses, we found that this stramenopile species has the gene encoding CLS_pld rather than CLS_cap, which stramenopiles exclusively possess according to Tian et al. (2012). Stramenopiles are a huge monophyletic assemblage of eukaryotes comprising Ochrophyta (composed of a wide variety of photosynthetic lineages including diatoms and brown algae) and two heterotrophic groups, Bigyra (including bicosoecids, labyrinthulomycetes, placidids, and

Blastocystis) and Pseudofungi (composed of oomycetes, hyphochytriomycetes, and *Developayella*) (Cavalier-Smith and Chao, 2006). Nevertheless, Tian et al. (2012) surveyed the presence or absence of CLS_cap/CLS_pld in a very limited number of phylogenetically narrow representatives of stramenopiles (five species of oomycetes, two species of diatom, one species of brown alga, and *Blastocystis*). Therefore, it remained uncertain whether many other stramenopile lineages contain CLS_cap or CLS_pld and whether only *Cantina* exceptionally possesses CLS_pld among stramenopiles.

In the present study, using genome or transcriptome data, we examined the presence or absence of the two types of CL biosynthesis enzyme, CLS_cap and CLS_pld, together with the CL maturation (remodeling) enzymes, CLD, iPLA2, TAZ, and ALCAT1, in stramenopile species more comprehensively in terms of phylogeny than in the case of Tian et al. (2012). Consequently, several species of stramenopiles other than *Cantina* were also found to have CLS_pld. We further traced the evolution of CLS_cap/CLS_pld within the radiation of stramenopiles based on the organismal phylogeny of this large eukaryotic group as resolved by large-scale phylogenetic analyses.

2. Materials and Methods

2.1. Cultures

The strains *Cafeteria roenbergensis* NIES1012, *Wobblia lunata* NIES1015, and *Developayella elegans* NIES1388 were purchased from the Microbial Culture Collection of the National Institute for Environmental Studies (NIES, 16-2 Onogawa, Tsukuba,

Ibaraki 305-8506, Japan). These cells were grown according to the instructions from the NIES.

2.2. Analyses of RNA-seq data from cultured strains

Cells of *C. roenbergensis*, *W. lunata*, and *D. elegans* were harvested by centrifugation at 4,170 g for 60 min at 4°C. Total RNA was isolated from the harvested cells using TRIzol reagent (Life Technologies, USA). Construction of cDNA libraries and paired-end sequencing with Illumina HiSeq 2000 (100 bp per read) were performed by Hokkaido System Science Co., Ltd. Raw sequencing reads were deposited in the Sequence Read Archive under the accession numbers DRX044597–DRX044599. 154–158 million raw sequencing read data were filtered using TRIMMOMATIC software version 0.30 (Lohse et al., 2012) to remove adapter sequences and low-quality bases. Filtered sequences were then assembled into 23,271–49,186 transcript contigs using the TRINITY package (release 2013-02-25) (Grabherr et al., 2011).

2.3. Sequence data collection

In addition to RNA-seq datasets from *C. roenbergensis*, *W. lunata*, and *D. elegans* generated in this study and from *Cantina marsupialis* generated in Noguchi et al. (2015), publicly available genome or transcriptome datasets from 80 taxa of eukaryotes (including 55 taxa of stramenopiles) were downloaded to find orthologous sequences for phylogenomic analyses and/or to identify the genes encoding enzymes for cardiolipin (CL) biosynthesis and maturation. Taxon names for which sequence datasets were

collected are listed in Table S1.

2.4. Phylogenomic analyses

Genome/transcriptome data as mentioned above were used as inputs for an in-house pipeline, described below, for the creation of single protein datasets and, subsequently, the phylogenomic data matrix. The organismal data were individually screened for orthologues using either blastp or tblastn, depending on the data type, with the reference orthologue sequences used as queries in BLASTMONKEY from the Barrel-o-Monkeys toolkit (<http://rogerlab.biochem.dal.ca>). If the sequence dataset was nucleotides, then the tblastn hits were translated to amino acid residues. Blastp was then used to screen these putative orthologues against the OrthoMCL database, and the output for each gene from each organism was compared against a dictionary of orthologous OrthoMCL IDs. Those putative orthologues that did not match orthologous IDs were designated as paralogues and removed. The remaining orthologues from each organism were combined and aligned using MAFFT-LINSI (Katoh and Standley, 2013). Ambiguously aligned positions were identified and removed using Block Mapping and Gathering with Entropy (BMGE) (Criscuolo and Gribaldo, 2010). For each individual protein alignment, maximum-likelihood (ML) trees were inferred in RAxML v7.2.6 (Stamatakis, 2006) using an LG model with four categories of among-site rate variation (LG + Γ 4), with 10 ML tree searches and 100 ML bootstrap replicates. To test for undetected paralogy or contaminants, we constructed a consensus tree (ConTree) representing phylogenetic groupings of well-established eukaryotic clades (Brown et al., 2013). The resulting

individual protein trees that placed taxa in conflicting positions relative to the ConTree with more than 70% ML bootstrap support, with a zero-branch length, or with extremely long branches were checked manually. All problematic sequences identified using these methods were removed from the dataset. The resulting protein alignments were then re-trimmed for ambiguously aligned positions using BMGE and concatenated into the separate supermatrix with 72,013 sites (245 proteins) of 84 taxa using alvert.py from the Barrel-o-Monkey's toolkit. The ML tree for the phylogenomic matrix was estimated from 60 independent searches using RAxML under the LG model with four categories of among-site rate variation and an empirical amino acid frequency model (LG + Γ 4 + F), selected by likelihood-ratio tests and the Akaike information criterion. Topological support was assessed using 1,000 RAxML bootstrap replicates. Because RAxML is not equipped with models that account for heterogeneous site-specific features of sequence evolution, we employed C-series models in IQ-TREE v1.3.3 (Nguyen et al., 2015) on the phylogenomic dataset. The best-fitting model available under ML analyses that we were capable of running with computational constraints was LG + Γ 4 + C10 + F with class weights optimized from the dataset using the exchangeabilities from the LG Q-Matrix (LG + Γ 4 + FMIX {empirical, C10pi1-C10pi10}) (Wang et al., 2014). Branch support was estimated from 1,000 ultrafast bootstrap replicates in IQ-TREE.

Evolutionary rates of change associated with each site of the phylogenomic matrix were estimated using DIST_EST (Susko et al., 2003) with the LG + Γ 4 model of evolution, using discrete gamma probability estimation (Susko et al., 2003). Using a custom python script, sites were removed from the phylogenomic matrix from the highest

(fastest) to the lowest (slowest) site rate in 3,000 site increments to exhaustion. Each incremental dataset was analyzed using 1,000 ultrafast bootstrap replicates in IQ-TREE under the LG + Γ 4 + F model of evolution. Further, a bootstrapped ML tree was inferred from the dataset that contained 45,013 of the original sites (approximately 63%), as this site-removal step showed maximal support in bootstrap values of nodes of interest in Fig. S1. These analyses were conducted under LG + Γ 4 + C10 + F model in IQ-TREE and under LG + Γ 4 + F in RAxML, each with 1,000 bootstrap replicates.

2.5. Identification of genes encoding proteins associated with CL biosynthesis and maturation

To screen homologues of two types of CL synthase (CLS_cap and CLS_pld), CL-specific phospholipase (CLD), calcium-independent phospholipase A2 (iPLA2), CoA-independent tafazzin (TAZ), and acylCoA:lysocardiolipin acyltransferase 1 (ALCAT1) from stramenopiles, blastp or tblastn was conducted against genome or transcriptome datasets from 59 stramenopile taxa using CLS_cap from *Phytophthora infestans* T30-4 (GenBank accession number EEY65382), CLS_pld from *Theileria parva* (EAN31681), CLD from *Ectocarpus siliculosus* (CBN79068), iPLA2 from *E. siliculosus* (CBN77084), TAZ from *P. infestans* T30-4 (XP_002904711), and ALCAT1 from *E. siliculosus* (CBJ33474) as queries with default settings. The top three retrieved hits for each protein from each taxon were further subjected to blastp homology searches against the Kyoto Encyclopedia of Genes and Genomes (KEGG) (<http://www.genome.jp/kegg>) database to remove possible paralogues of individual proteins. Among the sequences

1 surveyed against the KEGG database, those assigned to K numbers, K08744,
2 K06131/K06132, K13535, K16343, K13511, and K13513, were regarded as orthologues
3 of CLS_cap, CLS_pld, CLD, iPLA2, TAZ, and ALCAT1, respectively. Among these
4 screened orthologues, those with top hits that were bacterial homologues were potentially
5 derived from bacteria co-cultured with stramenopile species investigated in this study.
6 Thus, we considered only orthologues for which top hits were eukaryotic homologues as
7 originating from stramenopile species and utilized these in subsequent phylogenetic
8 analyses.

9 10 *2.6. Phylogenetic analyses of proteins for CL biosynthesis and maturation*

11 We conducted phylogenetic analyses of six proteins, CLS_cap, CLS_pld, CLD, iPLA2,
12 TAZ, and ALCAT1. The deduced amino acid sequences of these six proteins from
13 stramenopiles identified in this study were separately aligned with the corresponding
14 sequences from phylogenetically diverse organisms using MAFFT. The alignments were
15 inspected by eye and manually edited. We then excluded ambiguously aligned sites from
16 the datasets prior to phylogenetic analyses. The analyzed datasets had the following
17 dimensions: CLS_cap, 81 taxa, 141 sites; CLS_pld, 42 taxa, 490 sites; CLD, 43 taxa, 306
18 sites; iPLA2, 74 taxa, 236 sites; TAZ, 58 taxa, 272 sites; and ALCAT1, 105 taxa, 229 sites.
19 The alignment data are available on request to the corresponding author. For each
20 single-gene dataset, the ML phylogenetic tree and corresponding bootstrap support
21 values (1000 replicates) were calculated using RAxML. The ML tree was selected from
22 20 heuristic tree searches initiated from randomized parsimony starting trees. In ML

bootstrap analyses, a single tree search per replicate was performed. For the datasets of CLS_pld and CLS_cap, Bayesian analyses were also performed using MrBayes version 3.2.3. (Ronquist and Huelsenbeck, 2003). Six parallel Metropolis coupled Markov chain Monte Carlo (MCMCMC) runs, each consisting of three heated and one cold chains with default chain temperatures, were run for 1,000,000 generations. Log-likelihood scores and trees with branch lengths were sampled every 1,000 generations. The first 250,000 generations were excluded as burn-in, and the remaining trees were summarized to obtain Bayesian posterior probabilities. Convergence of parallel MCMCMC runs was judged by the average standard deviation of split frequencies (ASDSF). For both ML and Bayesian analyses, LG + Γ 4 or LG + Γ 4 + F model, which was selected as the most appropriate model with Aminosan (Tanabe, 2011), was applied.

3. Results

3.1. Phylogeny of stramenopiles

The phylogenetic tree based on 245 protein sequences from 84 eukaryote taxa (including 59 stramenopile taxa) is shown in Fig. 1. In this tree, Ochrophyta comprising 10 photosynthetic classes showed monophyly with 100% bootstrap probability (BP) of both RAxML and IQ-TREE. In the Ochrophyta clade, the three lineages composed of Pelagophyceae and Dictyochophyceae, of Bolidophyceae and Bacillariophyceae, and of Raphidophyceae, Xanthophyceae, and Phaeophyceae each formed a monophyly with 100% BP of both RAxML and IQ-TREE. The lineage of Pseudofungi composed of 17 oomycete taxa and *Developayella elegans* (Bigyromonadea) was monophyletic with 97%

BP of RAxML and 99% BP of IQ-TREE. The sister relationship between the lineages of Ochrophyta and Pseudofungi was supported with 100% BP of both RAxML and IQ-TREE. Four taxa of Labyrinthulomycetes formed a monophyletic clade with 100% BP of both RAxML and IQ-TREE, and this lineage was nested with the Ochrophyta/Pseudofungi clade with 80% BP of RAxML and 79% BP of IQ-TREE (note that this support increased up to almost 100% BP when excluding rapidly evolving sites from the dataset). A subset of *Bigyra*, *Cantina marsupialis*, *Wobblia lunata*, *Blastocystis hominis*, and two species of *Cafeteria* formed a clade with 100% BP of both RAxML and IQ-TREE, and this clade branched most basally in the stramenopile radiation. In this basal lineage, the anaerobic species *B. hominis* was grouped with *Cafeteria* sp. Caron Lab Isolate and *W. lunata* with 98% BP of RAxML and 100% BP of IQ-TREE, while the anaerobic species *C. marsupialis* was clustered with *Cafeteria roenbergensis* with 98% BP of RAxML and 100% BP of IQ-TREE. Two species of *Cafeteria* were not related in this phylogeny. Either of them may be misidentified based on inadequate morphological assessment.

We also evaluated the impact of the fastest evolving sites in the 72,013 amino acid supermatrix. These sites are theoretically the noisiest and are most prone to contribute to systematic error in the analysis, which may potentially yield artifactual results. To examine their impact on the phylogeny, sites were removed in a stepwise fashion from fastest to slowest evolving in 3,000 site increments and maximum likelihood (ML)-bootstrap analyses were performed under LG + Γ 4 + F for resulting alignments. The bootstrap values of nodes of interest are plotted in Fig. S1. This illustrates that

support for the larger clades of interest in stramenopiles remain stable when fast-evolving sites are removed. For example, the ML tree topology for the step 8 with 27,000 of the fastest evolving site removed (Fig. S2) was virtually identical to that of Fig. 1 (without fast site removal). Therefore, it is probable that the fast-evolving sites did not cause any significant artifact in our analyses.

3.2. Presence/absence of the genes encoding proteins associated with CL biosynthesis and maturation

The presence or absence of CLS_cap, CLS_pld, CLD, iPLA2, TAZ, and ALCAT1 in 59 stramenopile taxa examined in this study is summarized in Table S2, and the distribution of stramenopile CLS_cap/CLS_pld homologues is mapped on the organismal phylogenetic tree (Fig. 1). Only the CLS_cap homologue was identified in 41 taxa that are phylogenetically broad in the stramenopile lineage. However, only the CLS_pld homologue was detected in *C. marsupialis*, and both homologues of CLS_cap and CLS_pld were retrieved from *Cafeteria* sp. Caron Lab Isolate and *W. lunata*. Neither CLS_cap nor CLS_pld was recovered from the remaining 15 taxa. The homologues of the CL deacylation enzymes CLD and/or iPLA2 (beta or gamma) were detected in 44 taxa, while these two deacylation enzyme homologues were not recovered from the remaining 15 taxa. The homologues of the CL reacylation enzymes TAZ and/or ALCAT1 were found in 48 taxa, but these reacylation enzyme homologues were not retrieved from the remaining 11 taxa. A significant fraction of the homologues in question from stramenopiles was missing. As almost all datasets surveyed in this study were a

transcriptome or “draft” (incomplete) genome, we potentially missed the target sequences. Also, the absence of the enzymes associated with deacylation and/or reacylation of CL may mean that the CL maturation (remodeling) process is not necessarily indispensable for all stramenopiles.

3.3. Phylogenies of proteins for CL biosynthesis and maturation

In CLS_cap (Fig. 2) and CLS_pld (Fig. 3) trees reconstructed in this study, the monophyly of eukaryotic homologues was supported with 87% RAxML BP/1.00 Bayesian posterior probability (PP) and 100% RAxML BP/1.00 Bayesian PP, respectively. Stramenopile homologues of both CLS_cap and CLS_pld did not form monophyletic lineages in the respective eukaryotic clades, but such topology was not statistically supported. In the diatoms *Fragilariopsis cylindrus* and *Phaeodactylum tricornutum*, there were duplicates of the CLS_cap genes branched to each other, suggesting that gene duplication occurred at some point in the evolutionary lineages leading to these species. *W. lunata* was also found to contain two copies of the CLS_pld genes, but these copies were not related to each other in the eukaryotic radiation. The CLS_pld copies of *W. lunata* were possibly generated by gene duplication at the latest point before the divergence of *W. lunata* and *Cafeteria* sp. Caron Lab Isolate. In the phylogenies of all four CL maturation enzymes, CLD, iPLA2, TAZ, and ALCAT1 (Figs. S1–S4), deep phylogenetic relationships could not be resolved as in the case of Tian et al. (2012) and the phylogenetic status of the respective stramenopile homologues was not clear, making it difficult to infer the evolutionary processes of CL maturation enzymes

from stramenopiles.

4. Discussion

4.1. Deep phylogenetic relationships in stramenopiles

In contrast to previous phylogenetic analyses based on the small-subunit ribosomal RNA gene (e.g., Massana et al., 2014; Yubuki et al., 2015) and seven genes (Riisberg et al., 2009), our phylogenomic analyses based on 245 protein sequences successfully resolved deep relationships of stramenopiles as follows: 1) the monophyly of Pseudofungi composed of oomycetes and *Developyella* was well supported (although another Pseudofungi lineage, hypochytridiomycetes, was not included in our analyses). 2) The lineage of Labyrinthulomycetes, a member of Bigyra, was sister to the large clade comprising Ochrophyta and Pseudofungi. 3) Other members of Bigyra including bicosoecids (*Cafeteria* spp.), placidids (*Wobblia lunata*), *Cantina marsupialis*, and *Blastocystis hominis* formed a robust clade, and this clade diverged the earliest in the stramenopile radiation. 4) In this basal clade, two anaerobic protists, *Blastocystis* and *Cantina*, were not closely related to each other. Considering the second and third findings, the heterotrophic stramenopile group Bigyra is unlikely to be a natural assemblage, although Cavalier-Smith and Chao (2006) argued that Opalozoa (including *Blastocystis*), Bicoecia (including bicosoecids and placidids), and Sagenista (including labyrinthulids) are classified into this group based on single gene phylogenetic analyses. The fourth finding is also remarkable. Noguchi et al. (2015) reported that several enzymes involved in anaerobic pyruvate metabolism (such as ADP-forming acetyl-CoA synthetase and

1 acetate:succinate CoA-transferase) occur solely in *Blastocystis* and *Cantina* among
2 stramenopiles investigated and that the respective homologues from these two distantly
3 related anaerobic stramenopiles are very closely related to each other, suggesting
4 complex evolutionary scenarios of these enzymes required for anaerobic metabolism, as
5 previously discussed in Noguchi et al. (2015).

7 4.2. Evolution of CLS_cap and CLS_pld in the stramenopile radiation

8 Tian et al. (2012) reported that all nine representatives of stramenopiles they investigated
9 possessed CLS_cap but lacked CLS_pld, suggesting that CLS_pld was lost during the
10 course of evolution from the last eukaryotic common ancestor (LECA) having both
11 CLS_cap and CLS_pld to a common ancestor of all stramenopiles. Similarly, only
12 CLS_cap was detected in 41 stramenopile taxa in this study. However, we also found
13 several cases contradicting the theory of Tian et al. (2012): only CLS_pld was recovered
14 from *Cantina marsupialis*, and both CLS_cap and CLS_pld were found in *Cafeteria* sp.
15 Caron Lab Isolate and *Wobblia lunata*. It is not clear why *Cafeteria* sp. Caron lab isolate
16 and *Wobblia* retain both CLS_cap and CLS_pld. The redundant enzymes in these species
17 may have the same metabolic function. On the other hand, the possibility that either
18 CLS_cap or CLS_pld in these two species is involved in cardiolipin synthesis and that the
19 other is associated with an unknown metabolism cannot be ruled out. Notably, all these
20 CLS_pld-bearing species were members of the most basal lineage within the
21 stramenopile radiation. Based on these findings, new evolutionary scenarios regarding
22 CLS_cap/CLS_pld in the stramenopile group could be proposed, as shown in Fig. 4. The

common ancestor of all stramenopiles probably harbored both CLS_cap and CLS_pld. It is reasonable to assume that both types of CLS were inherited from the LECA, as suggested by Tian et al. (2012), because monophylies of these two types of CLS from a broad range of eukaryotes were respectively robust (Figs. 2 and 3). It is also highly likely that a common ancestor of the most basal stramenopile lineage retained both CLS_cap and CLS_pld and that *Cafeteria* sp. Caron Lab Isolate and *W. lunata* inherited these two CLSs. In contrast, *C. marsupialis* and *Cafeteria roenbergensis*/*B. hominis* may have lost CLS_cap and CLS_pld, respectively (only CLS_cap was detected in the latter). However, it cannot be fully ruled out that we failed to retrieve CLS_cap from *C. marsupialis* and CLS_pld from *C. roenbergensis*, because our analyses of these two species were based on transcriptome data rather than complete genome data. On the other hand, a common ancestor of Ochrophyta, Pseudofungi, and Labyrinthulomycetes (to the exclusion of the most basal stramenopile lineage) may have lost CLS_pld, as CLS_cap was exclusively detected in as many as 39 taxa belonging to these three lineages.

Tian et al. (2012) argued that each eukaryotic supergroup (such as Opisthokonta, Amoebozoa, Archaeplastida, or Excavata) has either CLS_cap or CLS_pld. Nevertheless, in this pioneering study, only CLS_cap was identified in stramenopiles, a subgroup of the supergroup SAR (composed of stramenopiles, Alveolata, and Rhizaria), while only CLS_pld was retrieved from alveolates. Considering this situation, Tian et al. (2012) doubted the close relationship between stramenopiles and alveolates. However, we newly found CLS_pld in several species of stramenopiles. In addition, there are multiple lines of reliable evidence that SAR is a monophyletic clade (e.g., Hackett et al.,

2007; Brown et al, 2013), and therefore we here propose the hypothesis that a common ancestor of stramenopiles and alveolates (or of SAR) possessed both CLS_{cap} and CLS_{pld}, with the subsequent loss of CLS_{cap} in the alveolate lineage. Thus, the doubt regarding the classification and relationship of stramenopiles and alveolates posed by Tian et al. (2012) could be dispelled.

4.3. CL in anaerobic protists

Based on biochemical evidence, *Trichomonas vaginalis* and *Giardia intestinalis*, representative anaerobic protists with functionally and morphologically degenerate mitochondria, so-called mitochondrion-related organelles (MROs), lack CL (Rosa et al., 2008; Guschina et al., 2009). In addition, Tian et al. (2012) reported that no genes encoding enzymes associated with CL biosynthesis and maturation were detected in the anaerobic MRO-bearing protists *Entamoeba* and microsporidians along with *Trichomonas* and *Giardia* using their genome or transcriptome data. On the other hand, the presence of CL in the anaerobic parasitic stramenopile *Blastocystis* and the anaerobic commensal ciliate inhabiting the cockroach hindgut *Nyctotherus ovalis*, both of which have MROs, was demonstrated (Keenan et al., 1992; Voncken et al., 2002). Furthermore, Stairs et al. (2014) identified the gene for CL biosynthesis (CLS_{cap}) in the free-living and MRO-bearing anaerobic protist *Pygsuia biforma* belonging to the eukaryotic lineage Breviatea. Similarly, in this study, the free-living anaerobic stramenopile *Cantina* with MROs was found to possess the gene for CL biosynthesis (CLS_{pld}). It should be noted that CL/CLS-lacking anaerobic protists (such as *Trichomonas*, *Giardia*, *Entamoeba*, and

microsporidians) are completely devoid of membrane-associated electron transport chain (ETC) complexes in their MROs, while MROs in CL/CLS-bearing ones (such as *Cantina*, *Blastocystis*, *Nyctotherus*, and *Pygсуia*) retain one or two ETC complexes (Stechmann et al., 2008; de Graaf et al., 2011; Stairs et al., 2014; Noguchi et al., 2015). Given that CL is responsible for the stabilization and activation of ETC complexes (e.g., Fry and Green, 1981; Eble et al., 1990; Robinson et al., 1990), the presence or absence of CL or its synthase in anaerobic protists may be reasonable.

5. Conclusions

Tian et al. (2012) pointed out that the respective major eukaryotic groups have one of the two types of cardiolipin synthase (CLS), CLS_{cap} or CLS_{pld}. They hypothesized that the last eukaryotic common ancestor (LECA) possessed both CLS_{cap} and CLS_{pld} and several differential losses of these two enzymes occurred at an early stage of eukaryotic evolution. For example, they argued that CLS_{cap} was exclusively found in the eukaryotic group stramenopiles, suggesting that CLS_{pld} was lost prior to the emergence of the common ancestor of all stramenopiles. However, in our study, one species with only CLS_{pld} (*Cantina marsupialis*) and two species with both CLS_{cap} and CLS_{pld} (*Cafeteria* sp. Caron Lab Isolate and *Wobblia lunata*) as well as many species with only CLS_{cap} were found among stramenopiles, and all three CLS_{pld}-bearing species were positioned in the most basal stramenopile lineage. Given these findings, it is highly likely that the common ancestor of all stramenopiles harbored both CLS_{cap} and CLS_{pld} and that these two enzymes were differentially lost in this group, as shown in Fig. 4. Our

present findings could also simultaneously support the hypothesis that both CLS_cap and CLS_pld existed in the LECA as originally proposed by Tian et al. (2012). The evolution of CLSs is not as simple even in a single eukaryotic clade, as suggested by the results of this study. Therefore, the evolutionary history of the enzymes in question in the eukaryotic domain as a whole is also possibly more complicated than previously thought, and major eukaryotic groups other than stramenopiles may have to be individually examined with comprehensive taxon sampling (indeed, the amoebozoan *Acanthamoeba* was found to possess both CLS_cap and CLS_pld through our analyses as shown in Figs. 2 and 3). In that case, a robust organismal phylogeny would be the key to trace the evolution of CLS_cap/CLS_pld precisely in each eukaryotic group.

Acknowledgments

We are grateful to Drs. Akinori Yabuki, Takao Yoshida, and Yoshihiro Fujiwara (JAMSTEC) and to Mr. Genki Ozawa (Kitasato University) for helpful discussion. This work was supported by a grant from the Japanese Society for the Promotion of Science awarded to K.T. (No. 15K07176) and the National Science Foundation Grant DEB 1456054 (<http://www.nsf.gov>) awarded to M.W.B. We thank the Mississippi State University's High Performance Computing Collaboratory for computational resources.

References

Adl, S.M., Simpson, A.G., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Le Gall,

1 L., Lynn, D.H., McManus, H., Mitchell, E.A., Mozley-Stanridge, S.E., Parfrey,
 2 L.W., Pawlowski, J., Rueckert, S., Shadwick, R.S., Schoch, C.L., Smirnov, A.,
 3 Spiegel, F.W., 2012. The revised classification of eukaryotes. *J. Eukaryot.*
 4 *Microbiol.* 59, 429–493.

5 Arias-Cartin, R., Grimaldi, S., Pommier, J., Lanciano, P., Schaefer, C., Arnoux, P.,
 6 Giordano, G., Guigliarelli, B., Magalon, A., 2011. Cardiolipin-based respiratory
 7 complex activation in bacteria. *Proc. Natl. Acad. Sci. USA* 108, 7781–7786.

8 Beranek, A., Rechberger, G., Knauer, H., Wolinski, H., Kohlwein, S.D., Leber, R., 2009.
 9 Identification of a cardiolipin-specific phospholipase encoded by the gene CLD1
 10 (YGR110W) in yeast. *J. Biol. Chem.* 284, 11572–11578.

11 Brown, M.W., Sharpe, S.C., Silberman, J.D., Heiss, A.A., Lang, B.F., Simpson, A.G.,
 12 Roger, A.J., 2013. Phylogenomics demonstrates that breviate flagellates are related
 13 to opisthokonts and apusomonads. *Proc. Biol. Sci.* 280, 20131755.

14 Cao, J.S., Liu, Y.F., Lockwood, J., Burn, P., Shi, Y.G., 2004. A novel
 15 cardiolipin-remodeling pathway revealed by a gene encoding an endoplasmic
 16 reticulum-associated acyl-CoA:lysocardiolipin acyltransferase (ALCAT1) in
 17 mouse. *J. Biol. Chem.* 279, 31727–31734.

18 Cavalier-Smith, T., Chao, E.E., 2006. Phylogeny and megasystematics of phagotrophic
 19 heterokonts (kingdom Chromista). *J. Mol. Evol.* 62, 388–420.

20 Chang, S.C., Heacock, P.N., Mileykovskaya, E., Voelker, D.R., Dowhan, W., 1998.
 21 Isolation and characterization of the gene (CLS1) encoding cardiolipin synthase in
 22 *Saccharomyces cerevisiae*. *J. Biol. Chem.* 273, 14933–14941.

1 Criscuolo, A., Gribaldo, S., 2010. BMGE (Block Mapping and Gathering with Entropy):
2 a new software for selection of phylogenetic informative regions from multiple
3 sequence alignments. BMC Evol. Biol. 10, 210.

4 de Graaf, R.M., Ricard, G., van Alen, T.A., Duarte, I., Dutilh, B.E., Burgtorf, C., Kuiper,
5 J.W., van der Staay, G.W., Tielens, A.G., Huynen, M.A., Hackstein, J.H., 2011.
6 The organellar genome and metabolic potential of the hydrogen-producing
7 mitochondrion of *Nyctotherus ovalis*. Mol. Biol. Evol. 28, 2379–2391.

8 Eble, K.S., Coleman, W.B., Hantgan, R.R., Cunningham, C.C., 1990. Tightly associated
9 cardiolipin in the bovine heart mitochondrial ATP synthase as analyzed by 31P
10 nuclear magnetic resonance spectroscopy. J. Biol. Chem. 265, 19434–19440.

11 Fry, M., Green, D.E., 1981. Cardiolipin requirement for electron transfer in complex I
12 and III of the mitochondrial respiratory chain. J. Biol. Chem. 256, 1874–1880.

13 Gold, V.A., Robson, A., Bao, H., Romantsov, T., Duong, F., Collinson, I., 2010. The
14 action of cardiolipin on the bacterial translocon. Proc. Natl. Acad. Sci. USA 107,
15 10044–10049.

16 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis,
17 X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N.,
18 Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K.,
19 Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq
20 data without a reference genome. Nat. Biotechnol. 29, 644–652.

21 Gu, Z.M., Valianpour, F., Chen, S.L., Vaz, F.M., Hakkaart, G.A., Wanders, R.J.A.,
22 Greenberg, M.L., 2004. Aberrant cardiolipin metabolism in the yeast *taz1* mutant: a

1 model for Barth syndrome. Mol. Microbiol. 51, 149–158.

2 Guschina, I.A., Harris, K.M., Maskrey, B., Goldberg, B., Lloyd, D., Harwood, J.L., 2009.

3 The microaerophilic flagellate, *Trichomonas vaginalis*, contains unusual acyl lipids

4 but no detectable cardiolipin. J. Eukaryot. Microbiol. 56, 52–57.

5 Hackett, J.D., Yoon, H.S., Li, S., Reyes-Prieto, A., Rümmele, S.E., Bhattacharya, D.,

6 2007. Phylogenomic analysis supports the monophyly of cryptophytes and

7 haptophytes and the association of rhizaria with chromalveolates. Mol. Biol. Evol.

8 24, 1702–1713.

9 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version

10 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

11 Keenan, T.W., Huang, C.M., Zierdt, C.H., 1992. Comparative analysis of lipid

12 composition in axenic strains of *Blastocystis hominis*. Comp. Biochem. Physiol. B

13 102, 611–615.

14 Koga, Y., Morii, H., 2007. Biosynthesis of ether-type polar lipids in archaea and

15 evolutionary considerations. Microbiol. Mol. Biol. Rev. 71, 97–120.

16 Lattanzio, V.M., Corcelli, A., Mascolo, G., Oren, A., 2002. Presence of two novel

17 cardiolipins in the halophilic archaeal community in the crystallizer brines from the

18 salterns of Margherita di Savoia (Italy) and Eilat (Israel). Extremophiles 6,

19 437–444.

20 Lobasso, S., Lopalco, P., Mascolo, G., Corcelli, A., 2008. Lipids of the ultra-thin square

21 halophilic archaeon *Haloquadratum walsbyi*. Archaea 2, 177–183.

22 Lohse, M., Bolger, A.M., Nagel, A., Fernie, A.R., Lunn, J.E., Stitt, M., Usadel, B., 2012.

RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 40, W622–627.

Malhotra, A., Edelman-Novemsky, I., Xu, Y., Plesken, H., Ma, J.P., Schlame, M., Ren, M.D., 2009. Role of calcium-independent phospholipase A(2) in the pathogenesis of Barth syndrome. *Proc. Natl. Acad. Sci. USA* 106, 2337–2341.

Massana, R., del Campo, J., Sieracki, M.E., Audic, S., Logares, R., 2014. Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *ISME J.* 8, 854–866.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274.

Nishijima, S., Asami, Y., Uetake, N., Yamagoe, S., Ohta, A., Shibuya, I., 1988. Disruption of the *Escherichia coli* *cls* gene responsible for cardiolipin synthesis. *J. Bacteriol.* 170, 775–780.

Noguchi, F., Shimamura, S., Nakayama, T., Yazaki, E., Yabuki, A., Hashimoto, T., Inagaki, Y., Fujikura, K., Takishita, K., 2015. Metabolic capacity of mitochondrion-related organelles in the free-living anaerobic stramenopile *Cantina marsupialis*. *Protist* 166, 534–550.

Riisberg, I., Orr, R.J., Kluge, R., Shalchian-Tabrizi, K., Bowers, H.A., Patil, V., Edvardsen, B., Jakobsen, K.S., 2009. Seven gene phylogeny of heterokonts. *Protist* 160, 191–204.

Robinson, N.C., Zborowski, J., Talbert, L.H., 1990. Cardiolipin-depleted bovine heart

1 cytochrome c oxidase: binding stoichiometry and affinity for cardiolipin
2 derivatives. *Biochemistry* 29, 8962–8969.

3 Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference
4 under mixed models. *Bioinformatics* 19, 1572–1574.

5 Rosa, I.D., Einicker-Lamas, M., Bernardo, R.R., Benchimol, M., 2008. Cardiolipin, a
6 lipid found in mitochondria, hydrogenosomes and bacteria was not detected in
7 *Giardia lamblia*. *Exp. Parasitol.* 120, 3215–3220.

8 Sakamoto, T., Inoue, T., Otomo, Y., Yokomori, N., Ohno, M., Arai, H., Nakagawa, Y.,
9 2012. Deficiency of cardiolipin synthase causes abnormal mitochondrial function
10 and morphology in germ cells of *Caenorhabditis elegans*. *J. Biol. Chem.* 287,
11 4590–4601.

12 Sandoval-Calderón, M., Geiger, O., Guan, Z., Barona-Gómez, F., Sohlenkamp, C., 2009.
13 A eukaryote-like cardiolipin synthase is present in *Streptomyces coelicolor* and in
14 most actinobacteria. *J. Biol. Chem.* 284, 17383–17390.

15 Schlame, M., 2008. Cardiolipin synthesis for the assembly of bacterial and mitochondrial
16 membranes. *J. Lipid Res.* 49, 1607–1620.

17 Schug, Z.T., Gottlieb, E., 2009. Cardiolipin acts as a mitochondrial signalling platform to
18 launch apoptosis. *Biochim. Biophys. Acta* 1788, 2022–2031.

19 Stairs, C.W., Eme, L., Brown, M.W., Mutsaers, C., Susko, E., Dellaire, G., Soanes, D.M.,
20 van der Giezen, M., Roger, A.J., 2014. A SUF Fe-S cluster biogenesis system in the
21 mitochondrion-related organelles of the anaerobic protist *Pygsuia*. *Curr. Biol.* 24,
22 1176–1186.

1 Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic
2 analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.

3 Stechmann, A., Hamblin, K., Pérez-Brocal, V., Gaston, D., Richmond, G.S., van der
4 Giezen, M., Clark, C.G., Roger, A.J., 2008. Organelles in *Blastocystis* that blur the
5 distinction between mitochondria and hydrogenosomes. *Curr. Biol.* 18, 580–585.

6 Susko, E., Field, C., Blouin, C., Roger, A.J., 2003. Estimation of rates-across-sites
7 distributions in phylogenetic substitution models. *Syst. Biol.* 52, 594–603.

8 Tanabe, A.S., 2011. Kakusan4 and Aminosan: two programs for comparing
9 nonpartitioned, proportional and separate models for combined molecular
10 phylogenetic analyses of multilocus sequence data. *Mol. Ecol. Resour.* 11,
11 914–921.

12 Tian, H.F., Feng, J.M., Wen, J.F., 2012. The evolution of cardiolipin biosynthesis and
13 maturation pathways and its implications for the evolution of eukaryotes. *BMC*
14 *Evol. Biol.* 12, 32.

15 Tuller, G., Hrastnik, C., Achleitner, G., Schiefthaler, U., Klein, F., Daum, G., 1998.
16 YDL142c encodes cardiolipin synthase (Cls1p) and is non-essential for aerobic
17 growth of *Saccharomyces cerevisiae*. *FEBS Lett.* 421, 15–18.

18 Voncken, F., Boxma, B., Tjaden, J., Akhmanova, A., Huynen, M., Verbeek, F., Tielens,
19 A.G., Haferkamp, I., Neuhaus, H.E., Vogels, G., Veenhuis, M., Hackstein, J.H.,
20 2002. Multiple origins of hydrogenosomes: functional and phylogenetic evidence
21 from the ADP/ATP carrier of the anaerobic chytrid *Neocallimastix* sp. *Mol.*
22 *Microbiol.* 44, 1441–1454.

1 Wang, H.C., Susko, E., Roger, A.J., 2014. An amino acid substitution-selection model
2 adjusts residue fitness to improve phylogenetic estimation. *Mol. Biol. Evol.* 31,
3 779–792.

4 White, D.C., Frerman, F.E., 1967. Extraction, characterization, and cellular localization
5 of the lipids of *Staphylococcus aureus*. *J. Bacteriol.* 94, 1854–1867.

6 Wriessnegger, T., Gubitz, G., Leitner, E., Ingolic, E., Cregg, J., de la Cruz B.J., Daum, G.,
7 2007. Lipid composition of peroxisomes from the yeast *Pichia pastoris* grown on
8 different carbon sources. *Biochim. Biophys. Acta* 1771, 455–461.

9 Yankovskaya, V., Horsefield, R., Törnroth, S., Luna-Chavez, C., Miyoshi, H., Léger, C.,
10 Byrne, B., Cecchini, G., Iwata, S., 2003. Architecture of succinate dehydrogenase
11 and reactive oxygen species generation. *Science* 299, 700–704.

12 Yubuki, N., Pánek, T., Yabuki, A., Čepička, I., Takishita, K., Inagaki, Y., Leander, B.S.,
13 2015. Morphological identities of two different marine stramenopile environmental
14 sequence clades: *Bicosoeca kenaiensis* (Hilliard, 1971) and *Cantina marsupialis*
15 (Larsen and Patterson, 1990) gen. nov., comb. nov. *J. Eukaryot. Microbiol.* 62,
16 532–542.

17 Zachman, D.K., Chicco, A.J., McCune, S.A., Murphy, R.C., Moore, R.L., Sparagna, G.C.,
18 2010. The role of calcium-independent phospholipase A(2) in cardiolipin
19 remodeling in the spontaneously hypertensive heart failure rat heart. *J. Lipid Res.*
20 51, 525–534.

21 Zinser, E., Sperka-Gottlieb, C.D., Fasch, E.V., Kohlwein, S.D., Paltauf, F., Daum, G.,
22 1991. Phospholipid synthesis and lipid composition of subcellular membranes in

the unicellular eukaryote *Saccharomyces cerevisiae*. J. Bacteriol. 173, 2026–2034.

Figure legends

Fig. 1. Maximum-likelihood (ML) phylogeny based on the 245-protein dataset from eukaryotic taxa including a broad range of 59 stramenopile taxa inferred by RAxML under the LG + Γ 4 + F model. ML bootstrap probabilities of RAxML (left) and IQ-TREE (right) are shown for bipartitions with support over 60%. Stramenopiles marked with a red “C” and a blue “P” represent taxa having CLS_cap and CLS_pld, respectively.

Fig. 2. Maximum-likelihood (ML) phylogeny of the cardiolipin synthase CLS_cap from a broad range of organisms inferred by RAxML under the LG + Γ 4 + F model. ML bootstrap probabilities are shown for bipartitions with support over 60%. Thick branches represent relationships with over 0.95 Bayesian posterior probabilities. Stramenopiles are shown in blue typeface.

Fig. 3. Maximum-likelihood (ML) phylogeny of the cardiolipin synthase CLS_pld from a broad range of organisms inferred by RAxML under the LG + Γ 4 model. ML bootstrap probabilities are shown for bipartitions with support over 60%. Thick branches represent relationships with over 0.95 Bayesian posterior probabilities. Stramenopiles are shown in blue typeface.

Fig. 4. Hypothetical evolutionary scenarios of CLS_cap and CLS_pld in the lineage of stramenopiles.

Figure 1

Figure 2

Figure 3

0.5 substitution/site

Figure 4