

Research article

Open Access

## Differential gene transfers and gene duplications in primary and secondary endosymbioses

Stefan Zauner<sup>2</sup>, Peter Lockhart<sup>1</sup>, Bettina Stoebe-Maier<sup>4</sup>, Paul Gilson<sup>3</sup>, Geoffrey I McFadden<sup>5</sup> and Uwe G Maier\*<sup>2</sup>

Address: <sup>1</sup>Allan Wilson Centre for Molecular Ecology and Evolution, Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand, <sup>2</sup>Cell Biology, Philipps-University Marburg, Karl-von-Frisch Str., 35032 Marburg, Germany, <sup>3</sup>Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3050, Australia, <sup>4</sup>Naturwissenschaften, An der Zahlbach 15, 35039 Marburg, Germany and <sup>5</sup>Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Victoria 3010, Australia

Email: Stefan Zauner - zauner@staff.uni-marburg.de; Peter Lockhart - p.j.lockhart@massey.ac.nz; Bettina Stoebe-Maier - bmaier@naturwissenschaften.de; Paul Gilson - gilson@wehi.edu.au; Geoffrey I McFadden - gim@unimelb.edu.au; Uwe G Maier\* - maier@staff.uni-marburg.de

\* Corresponding author

Published: 26 April 2006

Received: 22 December 2005

BMC Evolutionary Biology 2006, 6:38 doi:10.1186/1471-2148-6-38

Accepted: 26 April 2006

This article is available from: <http://www.biomedcentral.com/1471-2148/6/38>

© 2006 Zauner et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Most genes introduced into phototrophic eukaryotes during the process of endosymbiosis are either lost or relocated into the host nuclear genome. In contrast, *groEL* homologues are found in different genome compartments among phototrophic eukaryotes. Comparative sequence analyses of recently available genome data, have allowed us to reconstruct the evolutionary history of these genes and propose a hypothesis that explains the unusual genome distribution of *groEL* homologues.

**Results:** Our analyses indicate that while two distinct *groEL* genes were introduced into eukaryotes by a progenitor of plastids, these particular homologues have not been maintained in all evolutionary lineages. This is of significant interest, because two chaperone proteins always co-occur in oxygenic photosynthetic organisms. We infer strikingly different lineage specific processes of evolution involving deletion, duplication and targeting of *groEL* proteins.

**Conclusion:** The requirement of two *groEL* homologues for chaperon function in phototrophs has provided a constraint that has shaped convergent evolutionary scenarios in divergent evolutionary lineages. *GroEL* provides a general evolutionary model for studying gene transfers and convergent evolutionary processes among eukaryotic lineages.

### Background

Plastids, the solar powered energy factories of phototrophic eukaryotes, either translate mRNAs for their organelle-encoded genes or import nuclear-encoded proteins. In both cases, correct folding of proteins is managed by chaperones such as those of the GroEL family. These are an abundant class of chaperones, which are also found

in mitochondria, hydrogenosomes and prokaryotes [1,2]. Their importance and distribution has led to intensive investigation of their function, and has culminated in the 'molecular chaperon concept' [3,4], which has strongly influenced current understanding of protein folding and assembly.

Early genomics on the cyanobacterium *Synechocystis* sp. PCC 6803 and on plastid chromosomes of eukaryotes has highlighted a surprisingly varied distribution of genome locations for GroEL homologues amongst photosynthetic taxa [5,6]. *Synechocystis* sp. PCC 6803 harbours two different *groEL* genes, whereas only one is maintained in red algal plastomes and the plastid genome of the cyanelle. Genes encoding GroEL have not been located within the sequenced genomes of chloroplasts in green algae and land plants, but two nuclear homologs of *groEL*, the *cpn60s*, have been detected in the nuclear genome of *Chlamydomonas reinhardtii* and some land plants. In *Euglena gracilis*, no *groEL* gene has been identified [7]. The secondary endosymbionts of a cryptomonad (*Guillardia theta*) and also a diatom (*Odontella sinensis*) are known to encode a single *groEL* gene in their plastid, and it can be speculated that the presence of a single copy of *groEL* may indicate the ancestral state in a primary endosymbiont [8,9]. Interestingly, the nucleomorph genome of the cyp-tomonad *Guillardia theta* harbours a *groEL* homologue [6] and a *cpn60*-like homologue has recently also been found in the nucleomorph genome of another secondary endo-symbiont: the chlorarachniophyte *Bigeloviella natans* (Gilson & McFadden, unpublished). Additionally, *cpn60*-like genes have recently been discovered in the nuclear genomes of other photosynthetic organisms: including in a red alga (*Cyanidioschyzon merolae*), a diatom (*Thalassiosira pseudonana*), and *Plasmodium falciparum* (a parasitic organism that harbours a degenerate plastid). With the exception of *P. falciparum*, the co-occurrence of two *groEL* genes in the genomes of these recently sequenced organisms was predicted as necessary for maintaining chloroplast function [6]. We develop here a hypothesis for differential transfer and gene duplication that explains the distribution of *groEL* homologues amongst the multiple genomes of photosynthetic taxa. We discuss how these proteins may act as an important regulator for plastid functions.

## Results

### Substitution model selection

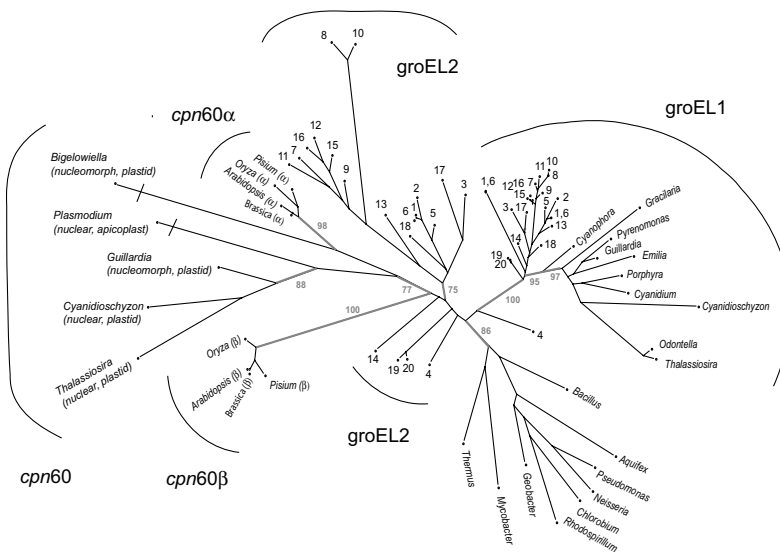
In all analyses, except analyses of the red/brown algal GroEL orthologues, an RtREV model, that accommodated positional rate heterogeneity in some form was the model selected as best by ProtTest under the AIC criterion. In the case of the GroEL orthologues in red/brown algae, we noted that small differences in the AIC criterion separated a variety of different models. When an RtREV model was assumed, and positional rate heterogeneity was approximated by either a constant proportion of variable sites or a discrete gamma distribution of rate classes, we noticed that the optimal estimates for these parameters varied among evolutionary lineages (i.e. different estimates were obtained for cyanobacterial GroEL1, cyanobacterial GroEL2, red algal/heterokont plastid GroEL1 like

sequences, and highly diverged Cpn60-like sequences). When a uniform rate distribution was assumed,  $p_{var}$  values ranged from 0.3 variable sites (cyanobacterial GroEL1) to 0.9 variable sites (highly diverged Cpn60-like sequences). When  $p_{var}$  was set to 1, alpha shape parameter values ranged from 0.2 (for cyanobacterial GroEL1) to 0.8 (for highly diverged Cpn60-like sequences). We suggest that this degree of variation in parameter estimates for phylogenetic grouping of anciently diverged sequences is likely to reflect lineage specific differences in structural and functional constraints [10-12]. As such it is potentially problematic for phylogenetic reconstruction, since parallel increases in proportions of variable sites in different evolutionary lineages can sometimes induce a form of long branch attraction [10-12].

### Evolutionary tree building

Optimal PhyML (protein maximum likelihood) trees showed similar topological relationships over a wide range of  $p_{var}$  values specified to accommodate positional rate heterogeneity. The robustness of phylogenetic relationships to sampling error was also found to be relatively stable, when evaluated using non-parametric bootstrapping (100 replicates). Figure 1 shows the optimal unrooted phylogenetic tree built assuming an RtREV model with positional rate heterogeneity modelled with a discrete gamma distribution ( $\alpha = 0.92$ ; 4 discrete rate categories).

Although, it contains more GroEL homologues than was available to Wastl et al. [6], our phylogenetic reconstruction here is nevertheless consistent with observations and inferences made by Wastl et al. [6]. Figure 1 shows that (a) the *groEL1* like genes from the plastid genome of red algae/heterokonts and cryptophytes are most closely related to the cyanobacterial *groEL1* genes. (b) Interestingly the cyanelle *groEL1* like homologue is somewhat intermediate between cyanobacterial and red algal/chromist *groEL1*-like sequences, a finding that may reflect the cyanobacterial-like nature of this plastid. (c) The non-photosynthetic eubacterial *groEL* sequences are arguably more similar to the *groEL2* sequences found in cyanobacteria. In any event, the *groEL2* orthologues in cyanobacteria are genetically more diverse than the *groEL1* sequences in the same taxa. These observations may suggest that *groEL2* orthologues represent an ancestral form of GroEL. However, our inference that structural/functional constraints differ amongst GroEL homologues, means that it is not possible to exclude other interpretations. (d) Assuming the root of the tree joins the branch leading to the non-photosynthetic taxa, Figure 1 places the *Gloeobacter* "A" and "B" sequences as the ancestral forms of GroEL homologues in photosynthetic taxa. (e) In the nuclear genomes of a diatom (*Thalassiosira pseudonana*), an alveolate (*Plasmodium falciparum*), the higher plants



**Figure 1**

Optimal PhyML tree, built assuming an RtREV + G( $\alpha = 0.92$ ) substitution model. Internal branches relevant to the discussion and which receive greater than 74% have been shown. Branch lengths are indicated. However, those subtending *Bigeloviella* and *Plasmodium* have been truncated. Cyanobacteria have been given number identifiers: [1] *Anabaena variabilis* 29413, [2] *Anabaena* sp. strain L-31, [3] *Crocosphaera watsonii* WH 8501, [4] *Gloeobacter violaceus* PCC 7421, [5] *Nostoc punctiforme* ATCC 29133, [6] *Anabaena* sp. PCC 7120, [7] *Prochlorococcus marinus* subsp. *marinus* str. CCMP1375, [8] *Prochlorococcus marinus* str. MIT 9312, [9] *Prochlorococcus marinus* str. MIT9313, [10] *Prochlorococcus pastoris* str. CCMP1986, [11] *Prochlorococcus marinus* str. NATL2A, [12] *Synechococcus* sp. CC9605, [13] *Synechococcus elongatus* 7942, [14] *Synechococcus vulcanus*, [15] *Synechococcus* sp. wh8102, [16] *Synechococcus* sp. CC9902, [17] *Synechocystis* PC 6803, [18] *Trichodesmium erythraeum* IMS101, [19] Cyanobacteria bacterium Yellowstone A-Prime, and [20] Cyanobacteria bacterium Yellowstone B-Prime.

and the nucleomorph genomes of a chlorarachniophyte (*Bigeloviella natans*), a cryptophyte (*Guillardia theta*) and a red algae (*Cyanidioschyzon merolae*) *groEL2* type sequences are found. In the case of higher plants, two forms of *groEL2* occur, and appear to represent forms of *groEL* duplicated from an ancestral *groEL2*-like sequence. Inferences concerning the origin of this duplication are potentially problematic. A relatively high bootstrap value in Figure 1 (77%) suggests that it may have occurred prior to the divergence of plastids. However, potential long branch attraction problems concerning the placement of  $\alpha$  and  $\beta$  Cpn60-like sequences from eukaryotes make this conclusion tentative. Gene duplication within the green lineage and differences in the functional/structural constraints of green  $\alpha$  and  $\beta$  Cpn60-like sequences might also explain the results observed. Although *Chlamydomonas reinhardtii* Cpn60  $\alpha$  and  $\beta$  sequences (AAA98642 and AAA98643) are not included in our phylogenetic analysis shown in Figure 1 (because their inclusion significantly reduced the alignment length, and increased phylogenetic uncertainty). However, it is clear from other phylogenetic analyses (unpublished) that this green alga contains the two forms of *groEL2* also present in higher plants.

**Discussion**

**Hypotheses of origin**

Our phylogenetic reconstruction suggests a complex pattern of genome transfers, losses and duplications in the evolution of *groEL* sequences from photosynthetic taxa. In the earliest cyanobacterial-like prokaryotes, the ancestral *groEL* sequence appears to have duplicated to form an "A" and a "B" type sequence that is still present today in the genome of *Gloeobacter* [14,15]. It appears that the "A" (*groEL2*-like) and "B" (more *groEL1*-like) forms have been inherited by most cyanobacteria, and also the endosymbiont(s) involved in primary plastid endosymbiosis. During the process of endosymbiosis, the *groEL2*-like orthologue has been transferred from the endosymbiont genome to the nuclear genome in heterokonts, red algae, green algae and high plants, and cryptophytes. We predict that this also will be true for haptophytes and glaucocystophytes. Less certain may be prediction of the pattern of evolution in peridinin-containing dinoflagellates, which in other respects appear highly distinctive in their evolution [16,17]. An interesting observation is that *Cyanophora paradoxa*, the molecular prototype of the glaucophytes, shows not only archaic features by encoding a *groES* in its

plastid genome, but also by containing a *groEL1* sequence, somewhat intermediate in sequence identity between cyanobacterial and plastid encoded *groEL1* genes. In the case of the chlorophyte lineage, it appears that it has been the *groEL1* form that has been lost, and in *Chlamydomonas* and higher plants, this has been replaced by a duplicated form of a *groEL2*-like sequence. This situation may also be true for the endosymbiont of the chlorarachniophyte *B. natans* (generally assumed to have been a green alga [16]), since phylogenetic reconstructions, including those with partial *Chlamydomonas* GroEL sequences (not shown), provide some support for *Bigeloviella* and *cpn60a* containing taxa being most closely related (e.g. Figure 1). Several investigations on the evolution of the *Plasmodium* apicoplast indicate a red algal origin for this organelle [18,19]. If so, then based on the generalisations stated above, one would expect that the *groEL/cpn60* distribution should be similar to the situation found in heterokonts. Additional duplications, of *groEL1* genes (in the presence of *groEL2*) have also occurred in some species of filamentous cyanobacteria and this observation is discussed in the following section.

#### GroEL proteins and protein folding

In all eukaryotes and prokaryotes that carry out oxygenic photosynthesis, and for which genome data is complete, two different forms of *groEL/cpn60* are known to exist. Thus, at least two copies of the gene or protein seem to be necessary for a complete chaperon function. Further, the importance of having two divergent forms of GroEL for protein folding is suggested from observations on GroEL sequences in *Synechocystis* sp. PCC 6803, where it has been observed that GroEL1 and GroEL2 respond differently to heat shock and light conditions [20,21]. An interesting speculation is that the number and genetic diversity of GroEL homologues reflects complexity of morphotype in cyanobacteria. This is suggested from comparison within filamentous cyanobacteria. Those strains that harbour three *GroEL* genes exhibit complex developmental stages (akinetes and hormogonia) whereas filamentous strains with a simpler morphotype, such as *Nostoc* sp. PCC 7120, harbour only two *GroEL* genes. Further study is needed to evaluate whether particular *groEL* homologues are specifically expressed in different developmental stages. Interestingly, *P. falciparum* encodes one *cpn60* gene, but no *groEL1* in the apicoplast genome. If phylogenetic inferences of a close relationship with red algae [18,19] are correct, then this observation may reflect relaxed constraints for protein folding for proteins of the apicoplast, and this speculation is also worth further investigation.

#### Conclusion

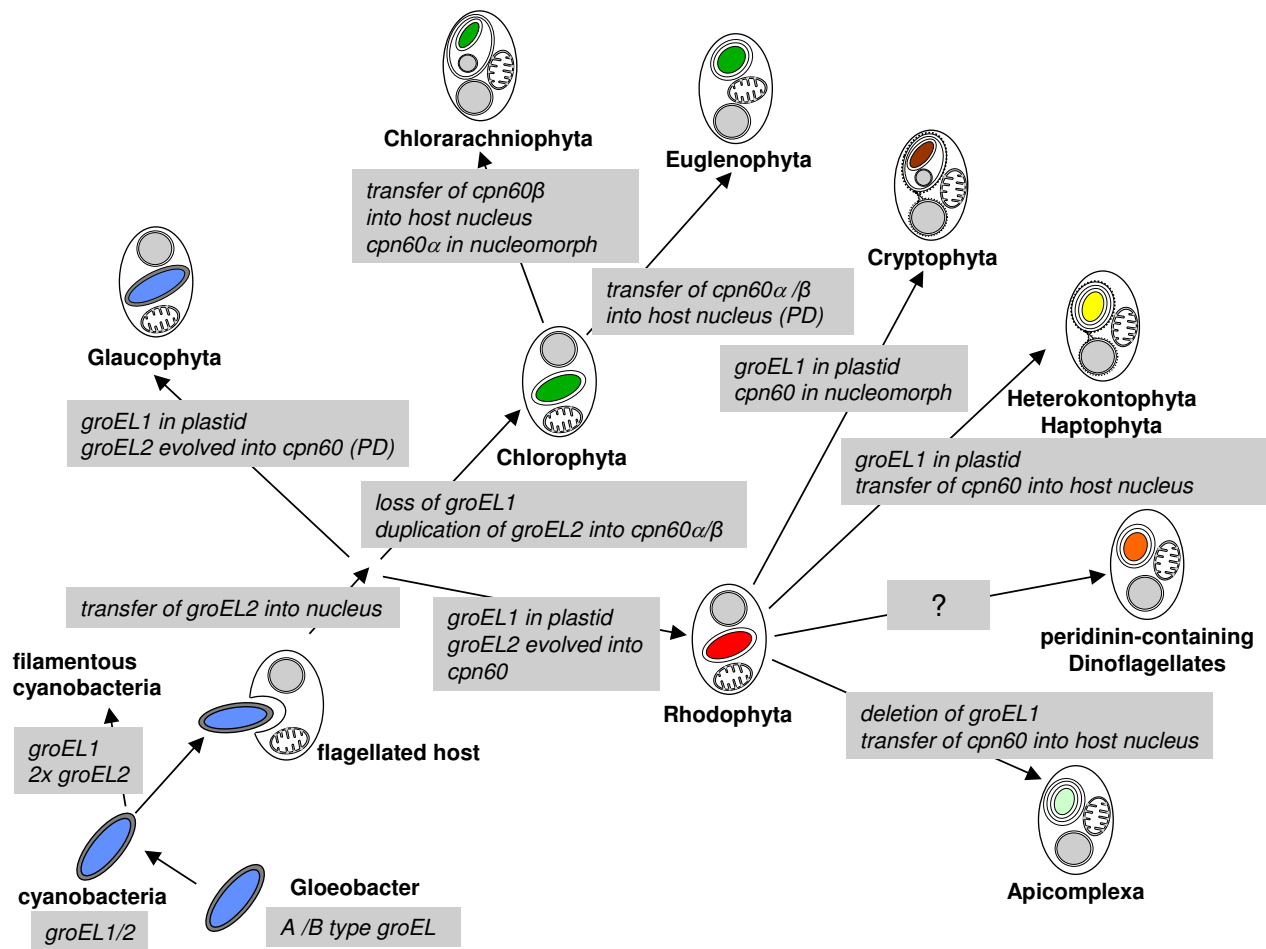
Although, phylogenetic reconstruction of individual gene histories is inherently problematic for anciently diverged taxa [5,10,11] phylogenetic reconstruction for *groEL*

homologues nevertheless provides a framework for developing understanding of genome-wide patterns of gene loss, relocation and multiple events of gene duplication. Our results presented here support and extend the hypothesis of *groEL/cpn60* evolution by Wastl et al. [6] which suggests a pattern of differential serial gene transfer and gene duplication.

#### Methods

##### Resources for sequences

Cyanobacterial *groEL* genes were retrieved from Genbank and the cpnDB chaperonin sequence database [22]. Two *groEL* genes were found to be present in *Synechocystis* sp. PCC 6803. One of these genes, termed *groEL1* (slr2076), is arranged in an operon together with *groES*, whereas the other, *groEL2* (sll0416), is not adjacent to a small subunit gene [23]. In comparing homologues from other cyanobacteria to those of *Synechocystis* sp. PCC 6803, we have adopted the terminology of "*groEL1*" and "*groEL 2*". All unicellular forms of cyanobacteria retrieved from database searches were found to contain two *GroEL* genes. In contrast, filamentous strains showed variation in the number of *groEL* genes. *Nostoc punctiforme* and *Anabaena variabilis* ATCC29413 harbour three *groEL* genes, whilst *Nostoc* sp. PCC 7120 with two *groEL* genes, was similar to that of unicellular forms. Genbank also provided us with several entries for *cpn60* from several land plants, *Chlamydomonas reinhardtii* and the nucleomorph of the cryptophyte *Guillardia theta*. In a recently finished genome project on the nucleomorph genome of the chlorarachniophyte *Bigeloviella natans* (Gilson & McFadden, unpublished), a further nucleomorph-encoded *cpn60* gene has been annotated. Two new genome projects on *Cyanidioschyzon merolae* [24] and *Thalassiosira pseudonana* [25] were also the source for additional nuclear-located *cpn60* genes. In PlasmoDB [26], the genome data base for *Plasmodium falciparum*, two different *cpn60* genes have been annotated in the nuclear genome [27], one encodes a mitochondrial, and the other an apicoplast targeted copy [28]. Additional, BLAST searches of NCBI were made against available plastid genomes using cyanobacterial *groEL* genes. Significant hits were obtained with the plastid genomes from red (*Gracilaria verrucosa*, *Cyanidium caldarium*, *Porphyra purpurea*, *Cyanidioschyzon merolae*) and secondary red plastids (*Odonetella sinensis*, *Guillardia theta*, *Pyrenomonas salina*, *Thalassiosira pseudonana*). These sequences, together with the nuclear, nucleomorph, plastid, and cyanobacterial sequences were aligned using the progressive alignment procedure implemented in CLUSTALX [29], edited to remove any ambiguously aligned regions, and conserved blocks of aligned residues (containing 495 amino acids) were used for phylogenetic analyses. The aligned data matrix and accession details are available from the authors on request.



**Figure 2**

Scheme depicting an evolutionary hypothesis that explains genome locations of *groEL* and *cpn60*. An ancestral eubacterial-like *groEL* duplicated in cyanobacteria to give two homologues: *groEL1* and *groEL2*. Both copies were inherited by phototrophic eukaryotes. One of the copies, *groEL1*, has been lost from the plastid genome in some lineages. *groEL2* was transferred into the cell nucleus and gave rise to *cpn60* in the case where *groEL1* is still maintained in the plastid genome, or into *cpn60α* and *cpn60β*, where *groEL1* has been deleted. For details see text. P.D. = predicted.

### Substitution model selection

The online version of ProtTest v1.2.6 [30,31], implementing the Akaike Information criterion (AIC) was used to select the most appropriate amino acid substitution models for tree building analyses. The cyanobacteria *GroEL1* and *GroEL2* datasets used for estimating parameters had identical taxon sampling (10 taxa: *Synechocystis* PC 6803, *Crocospaera watsonii* WH 8501, *Trichodesmium erythraeum* IMS101, *Synechococcus elongatus* 7942, *Synechococcus vulcanus*, *Synechococcus* sp. wh8102, *Prochlorococcus marinus* str. MIT9313, *Prochlorococcus marinus* str. CCMP1375, *Prochlorococcus pastoris* str. CCMP1986). Estimates were also made for bacteria (9 taxa: *Thermus thermophilus*, *Rhodospirillum rubrum*, *Neisseria meningitidis*, *Geobacter metallireducens*, *Pseudomonas aeruginosa*, *Mycobacterium avium*,

*Bacillus subtilis*, *Aquifex aeolicus* and *Chlorobium tepidum*; and for red/brown algal chloroplast located orthologues (8 taxa: *Guillardia theta*, *Pyrenomonas salina*, *Cyanidium caldarium*, *Porphyra purpurea*, *Gracilaria tenuistipitata* var. liui, *Odontella sinensis*, *Thalassiosira pseudonana* and *Cyanidioschyzon merolae*) and also for highly diverged Cpn60-like sequences (5 taxa: *Thalassiosira pseudonana*, *Cyanidioschyzon merolae*, *Plasmodium falciparum*, *Bigelowiella natans* and *Guillardia theta*).

### Evolutionary tree building

Trees were reconstructed from amino acid sequences using the windows version of PhyML [32]. Trees were built assuming an RTRev model and the optimal tree shown in Figure 1 displayed using SplitsTree4.0 [33]. The

robustness of phylogenetic reconstructions to variations in assumptions of positional rate heterogeneity was investigated by assuming (a) different proportions of variable sites ( $p_{var}$  range = 0.3–1.0) and (b) a discrete gamma distribution of rate classes and a range of alpha shape parameter values (0.2–1). Non-parametric bootstrap trees were analyzed to assess the significance of sampling variability.

### Authors' contributions

PG and GIM sequenced and analysed *cpn60* from *Bigelowiella natans*, and contributed to the manuscript. BSM has drawn Figure 2 and contributed to the manuscript. PJJ analysed the substitution properties of the data, constructed the phylogenetic trees and wrote together with UGM the manuscript. Data mining was done by SZ, who contributed to the manuscript as well. UGM has initiated the project. All authors read and approved the final manuscript.

### Acknowledgements

Our work is supported by the Deutsche Forschungsgemeinschaft (SFB TR I, TP A7), the Alexander von Humboldt Foundation, New Zealand Marsden Fund, the Australian Research Council, the National Health and Medical Research Council and the Howard Hughes Medical Institute.

### References

- Bui ET, Bradley PJ, Johnson PJ: **A common evolutionary origin for mitochondria and hydrogenosomes.** *Proc Natl Acad Sci U S A* 1996, **93**:9651-9656.
- Walter S: **Structure and function of the GroE chaperone.** *Cell Mol Life Sci* 2002, **59**:1589-1597.
- Ellis RJ, van der Vies SM: **Molecular chaperones.** *Annu Rev Biochem* 1991, **60**:321-347.
- Ellis RJ: **Protein folding: importance of the Anfinsen cage.** *Curr Biol* 2003, **13**:R881-R883.
- Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik KV: **Gene transfer to the nucleus and the evolution of chloroplasts.** *Nature* 1998, **393**:162-165.
- Wastl J, Fraunholz M, Zauner S, Douglas S, Maier UG: **Ancient gene duplication and differential gene flow in plastid lineages: the GroEL/Cpn60 example.** *J Mol Evol* 1999, **48**:112-117.
- Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E: **Complete sequence of Euglena gracilis chloroplast DNA.** *Nucleic Acids Res* 1993, **21**:3537-3544.
- Kowallik KV, Stoebe B, Schaffran I, Kroth-Pancic PG, Freier U: **The chloroplast genome of a chl a+c-containing alga, Odontella sinensis.** *Plant Mol Biol Reporter* 1995, **13**:336-342.
- Douglas SE, Penny SL: **The plastid genome of the cryptophyte alga, Guillardia theta: complete sequence and conserved synteny groups confirm its common ancestry with red algae.** *J Mol Evol* 1999, **48**:236-244.
- Lockhart PJ, Novis P, Milligan BG, Riden J, Rambaut A, Larkum AWD: **Heterotachy and tree building: a case study with plastids and eubacteria.** *Mol Biol Evol* 2005, **23**:40-45.
- Lockhart PJ, Steel MA: **A tale of two processes.** *Syst Biol* 2005, **54**:948-951.
- Guo Z, Stiller JW: **Comparative genomics and evolution of proteins associated with RNA polymerase II C-terminal domain.** *Mol Biol Evol* in press.
- Holland BR, Huber KT, Moulton V, Lockhart PJ: **Using consensus networks to visualize contradictory evidence for species phylogeny.** *Mol Biol Evol* 2004, **21**:1459-1461.
- Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, Tsuchiya T, Sasamoto S, Watanabe A, Kawashima K, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Takeuchi C, Yamada M, Tabata S: **Complete genome structure of *Gloeobacter violaceus* PCC a cyanobacterium that lacks thylakoids.** *DNA Res* 2003, **10**:137-145.
- Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, Tsuchiya T, Sasamoto S, Watanabe A, Kawashima K, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Takeuchi C, Yamada M, Tabata S: **Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids (supplement).** *DNA Res* 2003, **10**:181-201.
- Stoebe B, Maier UG: **One, two, three: nature's tool box for building plastids.** *Protoplasma* 2002, **219**:123-130.
- Keeling PJ: **Diversity and evolutionary history of plastids and their hosts.** *Amer J Bot* 2004, **91**:1481-1493.
- Foth BJ, McFadden GI: **The apicoplast: a plastid in Plasmodium falciparum and other Apicomplexan parasites.** *Int Rev Cytol* 2003, **224**:57-110.
- Coppin A, Varre JS, Lienard L, Dauvillee D, Guerardel Y, Soyer-Gobilard MO, Buleon A, Ball S, Tomavo S: **Evolution of plant-like crystalline storage polysaccharide in the protozoan parasite Toxoplasma gondii argues for a red alga ancestry.** *J Mol Evol* 2005, **60**:257-267.
- Glatz A, Horvath I, Varvasovszki V, Kovacs E, Torok Z, Vigh L: **Chaperonin genes of the Synechocystis PCC 6803 are differentially regulated under light-dark transition during heat stress.** *Biochem Biophys Res Commun* 1997, **239**:291-297.
- Kovacs E, van der Vies SM, Glatz A, Torok Z, Varvasovszki V, Horvath I, Vigh L: **The chaperonins of Synechocystis PCC 6803 differ in heat inducibility and chaperone activity.** *Biochem Biophys Res Commun* 2001, **289**:908-915.
- Hill JE, Penny SL, Crowell KG, Goh SH, Hemmingsen SM: **cpnDB: a chaperonin sequence database.** *Genome Res* 2004, **14**:1669-1675.
- Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hiroseawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S: **Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-encoding regions.** *DNA Res* 1996, **3**:109-136.
- Matsuzaki M, Misumi O, Shin IT, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, 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, Asakawa S, Takano H, Ohta N, Kuroiwa H, Tanaka K, Shimikzu N, Sugano S, Sato, Nozaki H, Ogasawara N, Kohara Y, Kuroiwa T: **Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D.** *Nature* 2004, **428**:653-657.
- Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavinia T, Goodstein D, Hadi MZ, Hellstein U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WW, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS: **The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism.** *Science* 2004, **306**:79-86.
- [http://www.plasmoDB.org].
- Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shalton SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Obornik M, Parker DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Borell B: **Genome sequence of the human malaria parasite Plasmodium falciparum.** *Nature* 2002, **419**:498-511.
- Sato S, Wilson RJ: **The use of DsRED in single- and dual-color fluorescence labeling of mitochondrial and plastid organelles in Plasmodium falciparum.** *Mol Biochem Parasitol* 2004, **134**:175-179.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL\_X windows interface: flexible strategies for mul-**

**Multiple sequence alignment aided by quality analysis tools.**  
*Nucleic Acids Res* 1997, **25**:4876-4882.

30. [<http://darwin.uvigo.es/>].
31. Abascal F, Zardoya R, Posada D: **ProtTest: Selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21**:2104-2105.  
[<http://atgc.lirmm.fr/phym1/>].
32. Huson DH, Bryant D: **Application of phylogenetic networks in evolutionary studies.** *Mol Biol Evol* 2006, **23**:254-267.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

