

## Phylogenetic analyses suggest lateral gene transfer from the mitochondrion to the apicoplast

Miroslav Oborník<sup>a,b,\*</sup>, Yves Van de Peer<sup>c</sup>, Václav Hypša<sup>a,b</sup>, Tancred Frickey<sup>e</sup>, Jan R. Šlapeta<sup>a,d</sup>, Axel Meyer<sup>e</sup>, Julius Lukeš<sup>a,b</sup>

<sup>a</sup>Institute of Parasitology, Czech Academy of Sciences, Branišovská 31, 370 05 České Budejovice, Czech Republic

<sup>b</sup>Faculty of Biological Sciences, University of South Bohemia, České Budejovice, Czech Republic

<sup>c</sup>Department of Molecular and Plant Genetics, Flanders Interuniversity Institute for Biotechnology, Ghent University, B-9000 Ghent, Belgium

<sup>d</sup>Department of Parasitology, University of Veterinary and Pharmaceutical Sciences, Brno, Czech Republic

<sup>e</sup>Faculty of Biology, University of Konstanz, Konstanz, Germany

Received 5 September 2001; received in revised form 12 December 2001; accepted 21 January 2002

Received by W.-H. Li

### Abstract

Apicomplexan protozoa contain a single mitochondrion and a multimembranous plastid-like organelle termed apicoplast. The size of the apicomplexan plastid genome is extremely small (35 kb) thus offering a limited number of genes for phylogenetic analysis. Moreover, the sequences of apicoplast genes are highly adenosine + thymidine-rich and rapidly evolving. Due to these facts, phylogenetic analyses based on different genes or the structure of the ribosomal operon show conflicting results and the evolutionary history of this exciting organelle remains unclear. Although it is evident that the apicoplast and its genome is plastid-derived, our detailed phylogenetic analysis of amino acid and nucleotide sequences of selected apicoplast ribosomal protein genes *rpl2*, *rpl14* and *rps12* show their possible mitochondrial origin. The affinity of apicoplast ribosomal proteins to their mitochondrial homologs is very stable and well supported. Based on our results we propose that apicoplasts might contain both plastid and mitochondrial genes, thus constituting a hybrid assembly. © 2002 Elsevier Science B.V. All rights reserved.

**Keywords:** Apicoplast; Ribosomal protein; Plastid; Mitochondrion; Hybrid genome; Gene transfer

### 1. Introduction

Apicomplexan protozoa contain two organellar genomes. Firstly, the ~35 kb circular element has been found in the avian parasite *Plasmodium lophurae* (Kilejian, 1975). Because of its clearly prokaryotic character it was concluded that it originated from the mitochondrion.

However, in 1989, the second apicomplexan extrachromosomal DNA was discovered (Vaidya et al., 1989). This ~6 kb linear molecule represents the smallest mitochondrial genome known (Lang et al., 1997), and codes for only three respiratory chain proteins and two truncated ribosomal RNAs (Vaidya et al., 1989). Since the existence of two different mitochondrial genomes within a single cell was quite implausible, the origin of the 35 kb DNA circle had begun to be intensively studied. Based on the close phylogenetic relationship between apicomplexans and dinoflagellates (Cavalier-Smith, 1993; Van de Peer et al., 1996) and the fact that the 35 kb circular molecule has a gene content similar to that of chloroplasts, hypothesis that it represents a highly reduced plastid genome started to emerge. The DNA circle was localized *in situ* to a multimembranous structure, previously described as a 'spherical body' (e.g. Hepler et al., 1966) or a 'multimembranous vesicle' (e.g. Schrével, 1971), identified recently as a secondary vestigial plastid, and coined the term apicoplast (McFadden et al., 1997; Köhler et al., 1997).

Abbreviations: aa, amino acid(s); AT, adenosine + thymidine; *clpC*, gene for caseinolytic protease C; D, distance (optimality criterion); kb, kilobase(s); LBA, long branch attraction; *lsu*, large subunit; ML, maximum likelihood (optimality criterion); MP, maximum parsimony (optimality criterion); mt, mitochondrial; nt, nucleotide(s); pl, plastid; *rbcL*, gene for ribulose-1,5-bisphosphate carboxylase, large subunit; *rp(s)*, gene for ribosomal protein(s); *rpl2*, gene for ribosomal protein L2; *rpl11*, gene for ribosomal protein L11; *rpl14*, gene for ribosomal protein L14; *rps8*, gene for ribosomal protein S8; *rps12*, gene for ribosomal protein S12; *rpoB*, gene for RNA polymerase, subunit B; *rpoC*, gene for RNA polymerase, subunit C; *rRNA*, ribosomal RNA; *ssu*, small subunit; *tRNA*, gene for transfer RNA; *tufA*, gene for elongation factor Tu; *ycf24*, gene for *ycf24* protein

\* Corresponding author.

E-mail address: obornik@paru.cas.cz (M. Oborník).

The ~35 kb DNA circle located in a plastid-like organelle represents the smallest plastid genome known so far. Besides a set of *tRNAs*, *rRNAs*, *rpoB*, *rpoC*, *tufA*, and ribosomal protein (*rp*) genes, it contains genes typical for plastids (*ycf24* and *clpC*) and several open reading frames of unknown function (McFadden et al., 1997; Köhler et al., 1997). Translation with prokaryotic features may be present in the apicoplast (Hopkins et al., 1999); however, most of its proteins appear to be nucleus-encoded and imported into the organelle, using an import secretory pathway (Waller et al., 2000).

The non-photosynthetic plastid in *Toxoplasma gondii* is surrounded by four membranes (Köhler et al., 1997); however, a detailed electron microscopy study of the plastid in *Plasmodium falciparum* showed three surrounding membranes with an outer and inner multimembranous vesicular complexes (Hopkins et al., 1999). Surprisingly, apicoplast is, during the entire life cycle, in close proximity, almost a physical contact, with the mitochondrion. At the same time, both organelles are present in a single copy per plasmodial cell (Hopkins et al., 1999).

Phylogenetic analyses based on the apicoplast *rRNAs* and *tufA* gene sequences further support the plastid origin of the 35 kb molecule (Köhler et al., 1997; Blanchard and Hicks, 1999; Zhang et al., 2000). The sequences of most apicoplast genes, including the *rp* genes (Blanchard and Hicks, 1999), suggest that these, and by inference the plastid, originate from a green alga (Köhler et al., 1997; Blanchard and Hicks, 1999). However, the structure of the ribosomal operon (McFadden et al., 1997; Blanchard and Hicks, 1999), as well as the *rRNA* genes (Zhang et al., 2000) most resembles those from a red alga. The exact origin of

the apicoplast remains, therefore, uncertain. Recent phylogenetic analysis based, apart from other coding genes, on concatenated *rp* genes showed their clustering with green plastids, but did not include mitochondrial sequences (Blanchard and Hicks, 1999). The addition of mitochondrial homologs into the analysis revealed a strong mitochondrial signal in the apicoplast *rp* genes. We propose a hypothesis that the apicoplast genome can be a mixture of plastid and mitochondrial genes.

## 2. Methods

### 2.1. Likelihood, parsimony and distance analyses

Amino acid (aa) sequences of three *rp* genes *rpl2*, *rpl14* and *rps12* from the apicoplasts of *Toxoplasma gondii* and *Plasmodium falciparum* and their homologs available from bacteria, plastids, and mitochondria (see Table 1 for accession numbers) were aligned at the aa level using MegAlign 4.0 (DNA Star) and the alignment was manually improved. Subsequently, aa sequences were backtranslated into their nucleotide sequences (nt) and both aa and nt alignments were used for phylogenetic analyses. Regions that could not be aligned unambiguously were removed from the alignment. Both the aa and nt sequence alignments are available from the authors upon request. Phylogenetic analyses were carried out using pairwise distance methods (D), maximum parsimony (MP), and maximum likelihood (ML) on the basis of a concatenated alignment of the three *rp* genes. Software packages used for phylogenetic analyses were TREECON (Van de Peer and De Wachter, 1994), PAUP\*

Table 1  
List of species and ribosomal proteins accession numbers used in the current analysis

Organism	Taxon	Accession no.
<i>Bacillus subtilis</i>	Proteobacteria	Z99104
<i>Escherichia coli</i>	Proteobacteria	U00096
<i>Haemophilus influenzae</i>	Proteobacteria	L42023
<i>Mycobacterium leprae</i>	Actinobacteria	Z98756; L40409
<i>Mycobacterium tuberculosis</i>	Actinobacteria	Z84395; AL021943
<i>Rickettsia prowazekii</i>	Proteobacteria	AJ235270; AJ235272
<i>Synechocystis</i> sp.	Cyanobacteria	D90905; D90913
<i>Thermotoga maritima</i>	Thermotogales	AE001798; E001799
<i>Treponema pallidum</i>	Spirochaetales	AE000520
<i>Acanthamoeba castellanii</i>	Mitochondrion	U12386
<i>Dictyostelium discoideum</i>	Mitochondrion	NC000895
<i>Paramecium aurelia</i>	Mitochondrion	NC001324
<i>Reclinomonas americana</i>	Mitochondrion	AF007261
<i>Tetrahymena pyriformis</i>	Mitochondrion	AF160864
<i>Chlorella vulgaris</i>	Green primary plastid	NC001865
<i>Euglena gracilis</i>	Green secondary plastid	X70810
<i>Guillardia theta</i>	Red secondary plastid	AF041468
<i>Nicotiana tabacum</i>	Green primary plastid	Z00044
<i>Odontella sinensis</i>	Red secondary plastid	Z67753
<i>Porphyra purpurea</i>	Red primary plastid	U38804
<i>Plasmodium falciparum</i>	Apicoplast	X57167
<i>Toxoplasma gondii</i>	Apicoplast	U87145

4.0v8 (Swofford, 1998), and TREE-PUZZLE (Strimmer and von Haeseler, 1996). Distance trees inferred from the aa sequences were computed on the basis of the Poisson and the Kimura distances as implemented in TREECON (Van de Peer and De Wachter, 1994). Distance trees inferred from nt sequences were computed using the LogDet-paralinear distances (Lockhart et al., 1994) as implemented in the PAUP\* program (Swofford, 1998). ML trees were computed taking into account the among-site rate variation with eight rate categories, and using different substitution probability matrices as implemented in TREE-PUZZLE (Strimmer and von Haeseler, 1996). To test the possibility that the similarity between the apicoplast and mitochondrial *rps* might be due to convergent evolution of an anomalous nucleotide composition after the loss of the photosynthetic function, rather than to the common ancestry, the *rpl2* and *rps12* sequences from a non-photosynthetic plastid of the euglenoid *Astasia longa* were included in separate analyses (for explanation see Section 4.3).

Inconsistencies in tree topologies may also be caused by differences in the nt or aa sequences composition. To test whether the attraction between apicoplast and mitochondrial genes is artificial and due to a bias in the aa sequence content, we constructed neighbor-joining trees from the aa LogDet distances (Penny et al., 1999), assuming different proportions of the sites being invariant.

## 2.2. Phylogenetic analysis with respect to the saturation of amino acids

Apart from the standard phylogenetic techniques for the inference of evolutionary trees, we also applied a new method to correct for saturation at the aa level. Recently, some of us developed the AsaturA program to estimate the amount of mutational saturation in aa sequences alignments (Van de Peer et al., 2002). Since saturation is most likely to occur at sites that are easily substituted, AsaturA discriminates between aa substitutions that have a high probability of occurrence and those with low probability of aa replacement. Therefore, all aa replacements are defined as either 'frequent' or 'rare' depending on their mutation probabilities, which can be found in a so-called substitution probability matrices (Van de Peer et al., 2002). In AsaturA, all substitution probabilities of a selected substitution probability matrix are sorted and a probability 'cutoff' value can be chosen above which aa replacements are classified as 'frequent'; aa replacements for which the substitution probability is lower than the chosen value are then classified as 'rare'. The program then plots, for each sequence pair, the number of observed frequent and rare aa replacements against their evolutionary distance. By changing the substitution probability 'cutoff' value, the number of aa substitutions classified as frequent or rare can be changed. As a result, the degree of saturation can be evaluated for different sets of positions, and a careful selection of the 'cutoff' value splits the original data set into a saturated and unsaturated

cluster. After identifying aa replacements that are likely to be saturated, evolutionary distances between sequences can be computed from the unsaturated sites only and evolutionary trees inferred by pairwise distance methods such as neighbor-joining (Van de Peer et al., 2002).

## 3. Results

### 3.1. Genes used for the analyses

We compared the apicoplast genes *rpl2*, *rpl14* and *rps12* of *P. falciparum* and *T. gondii* with their homologs from non-photosynthetic bacteria, cyanobacterium, green and red primary and secondary plastids, and mitochondria (Table 1). Other apicoplast genes coding for *rps* could not be analyzed, either because extreme divergence makes them unalignable, or because homologs are available only for a narrow set of organisms. Since the result of a single-gene approach is often biased due to the low number of phylogenetically informative positions (e.g. Baldauf et al., 2000), we used a concatenated alignment of the three above mentioned *rp* genes, unless mentioned otherwise. Since these *rp* genes are conservative in length, no indels specific for plastids or mitochondria were found.

### 3.2. Phylogenetic analyses using amino acid sequences

As expected, in all our phylogenetic analyses based on the concatenated aa sequence alignment using D, MP, or ML, the mitochondrial *rp* genes form a monophyletic lineage and share a common origin with the  $\alpha$ -proteobacterium *Rickettsia prowazekii* (Fig. 1). Within the mitochondrial cluster, the primitive mitochondrion of *Reclinomonas americana* (Lang et al., 1997) forms the earliest branch, while the plastid and cyanobacterial *rps* form a second monophyletic clade. Rather unexpectedly, the apicoplast *rp* sequences invariably appeared within the mitochondrial lineage, regardless of the phylogenetic method used. Furthermore, the apicoplast sequences cluster with those of the ciliates, which are known to be closely related to Apicomplexa (Baldauf et al., 2000; Van de Peer et al., 2000). Bootstrap analysis and the TREE-PUZZLE tests (Strimmer and von Haeseler, 1996) strongly support the monophyly of both plastid and mitochondrial clades, from which the mitochondrial clade contains both apicoplast sequences. The bootstrap and TREE-PUZZLE results are summarized in Table 2.

As shown in Fig. 1, both the mitochondrial and apicoplast sequences are rapidly evolving. To investigate whether the clustering of mitochondria and apicoplast sequences is an artifact due to the attraction of long branches (LBA), trees were constructed using only the most conserved parts of the sequence alignment and based on a dataset where most long branches were excluded. Nevertheless, even when only one mitochondrial sequence was included, apicoplasts and mitochondria clustered together. When tested using MP and D

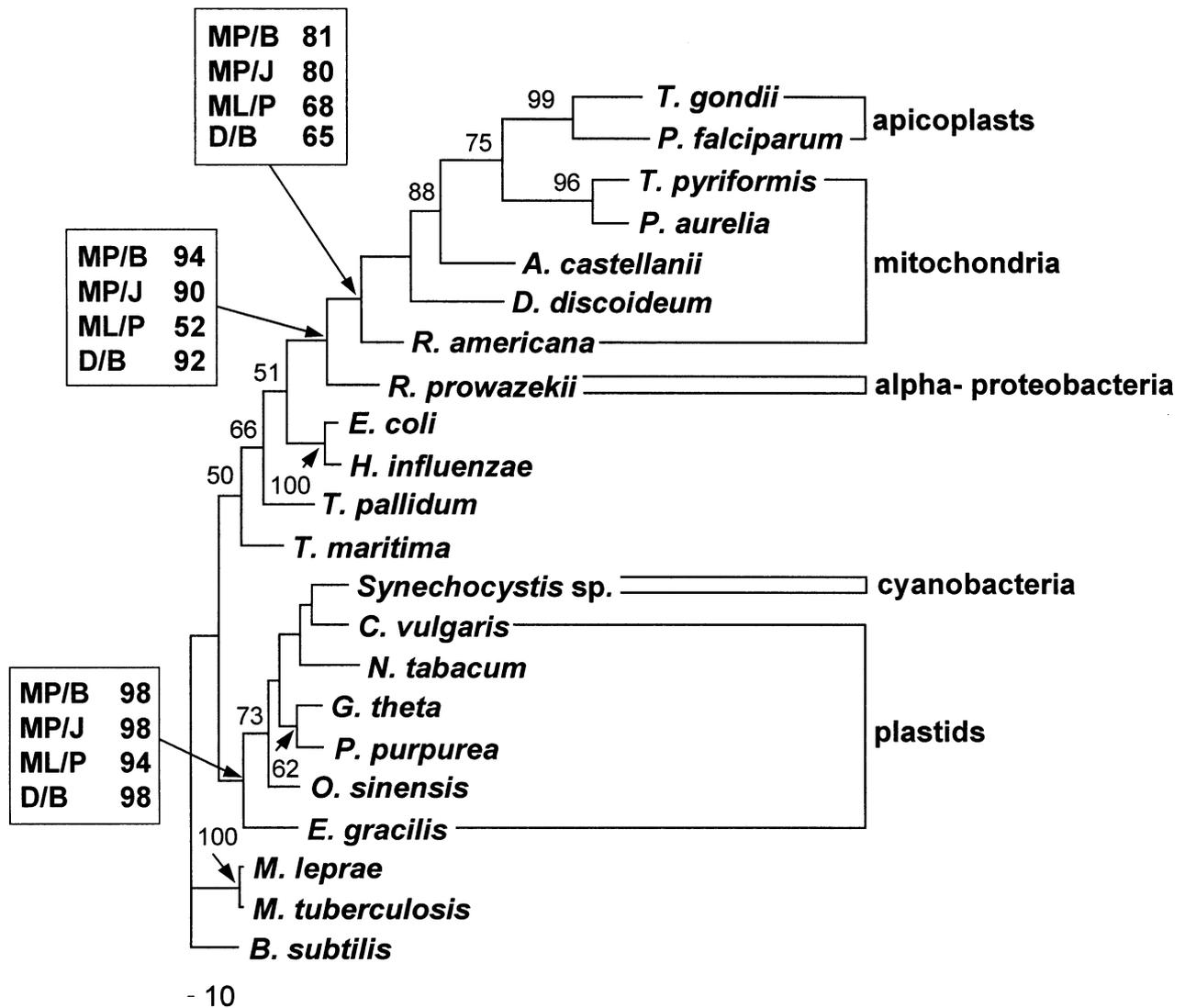


Fig. 1. Maximum parsimony tree inferred from concatenated alignment of *rpl2*, *rpl14*, and *rps12* aa sequences. Tree was calculated out of 513 characters, from which 40 were constant, 103 variable uninformative and 370 variable informative. The length of the tree = 2829 steps; CI = 0.6858; HI = 0.3142; CI (excluding uninformative characters) = 0.6594; RI = 0.4858; RC = 0.3332. Tree was rooted using *B. subtilis* sequences as an outgroup. Maximum parsimony bootstrap supports (1000 replicates) are indicated, support values of selected nodes using different optimality criteria are included. MP/B, maximum parsimony bootstraps; MP/J, maximum parsimony Jackknife monophyly indices; ML/P, maximum likelihood puzzle supports; D/B, distance bootstraps.

such clustering was well supported by the bootstrap analysis. The apicoplast clustering is unresolved when all mitochondria are excluded.

When the dataset containing *Astasia longa* plastid *rp* genes was analyzed, the genes from this non-photosynthetic plastid branched within the plastid lineage together with their homologs in the photosynthetically active relative *Euglena gracilis* (data not shown). This topology is stable regardless of the stringency of alignment or method used and provides further support for the mitochondrial signal of the analyzed apicoplast sequences.

Analysis based on the aa LogDet distances showed the affinity of the apicoplast *rp* genes to their mitochondrial homologs as well. However, despite the fact that the mitochondrial and apicoplast sequences clustered together, in

particular when larger fractions of invariant sites were removed, this was not supported by high TREE-PUZZLE values. Furthermore, the grouping of apicoplasts and mitochondria usually did not include *Rickettsia* or *Reclinomonas* while some other groupings were also peculiar (data not shown). However, this might not be too surprising, since good estimates require long sequences and even combining the three ribosomal proteins (a total of 425 positions) may not be enough to yield statistically valid estimates (Penny et al., 1999).

As shown in Fig. 2B, a considerable fraction of the aa replacements is saturated, which may be the cause of erroneous tree topologies. Therefore, a tree was constructed taking into account only the unsaturated fraction of the aa replacements (see Section 2.2 and Fig. 2) (Van de Peer et al.,

Table 2  
Nodal support of constructed trees<sup>a</sup>

Criterion	Node	aa	aa*	nt	nt*
Maximum parsimony	A	94/81	–	36/84	33/64
	B	98	–	27	92
Maximum likelihood	A	52/52	–	36/76	76/52
	B	94	–	38	95
Distance	A	92/65	98/99	59/80	72/51
	B	98	99	86	98

<sup>a</sup> Bootstrap support of the mitochondrial (A; left value excluding the alpha proteobacterium *Rickettsia*, the right value including *Rickettsia*) and plastid (B) cluster as inferred from concatenated ribosomal protein sequences. Trees were constructed based on the amino acid alignment (aa), the amino acid alignment with excluded saturated positions (aa\*; see Section 2.2 for details), the nucleotide alignment (nt), and the nucleotide alignment with third codon position excluded (nt\*). Distance nt trees were computed using LogDet. Trees constructed on the basis of aa\* were computed only using the neighbor-joining method (see Section 2.2). Bootstrap values were determined (out of 1000 replicates) for all trees except for the maximum-likelihood tree based on aa sequences, where quartet-puzzling support was calculated instead. In all trees apicoplast sequences appeared within the mitochondrial cluster (Figs. 1–4).

2002). In this bootstrapped tree, apicoplast sequences appeared again within the mitochondrial cluster. Furthermore, they cluster again with the ciliate mitochondrial sequences, and the monophyletic grouping of mitochondria and  $\alpha$ -proteobacteria (represented by *R. prowazekii*) is also highly supported.

### 3.3. Phylogenetic analyses using nucleotide sequences

We also backtranslated the aa sequence alignment into the nt sequence alignment. As with the aa sequences, a concatenated alignment of all three *rps* was constructed and used for further analysis. MP, ML, and D trees show topologies similar to those inferred from the aa sequences (Figs. 1–4); however, the bootstrap support for some trees is somewhat lower (Table 2). Since nucleotides in the third codon position are saturated (data not shown), we have also analyzed the alignment considering the first and second codon positions only. Overall, omitting the third codon positions results in the increase of bootstrap values, in particular for the plastid cluster (Table 2). The position of the apico-

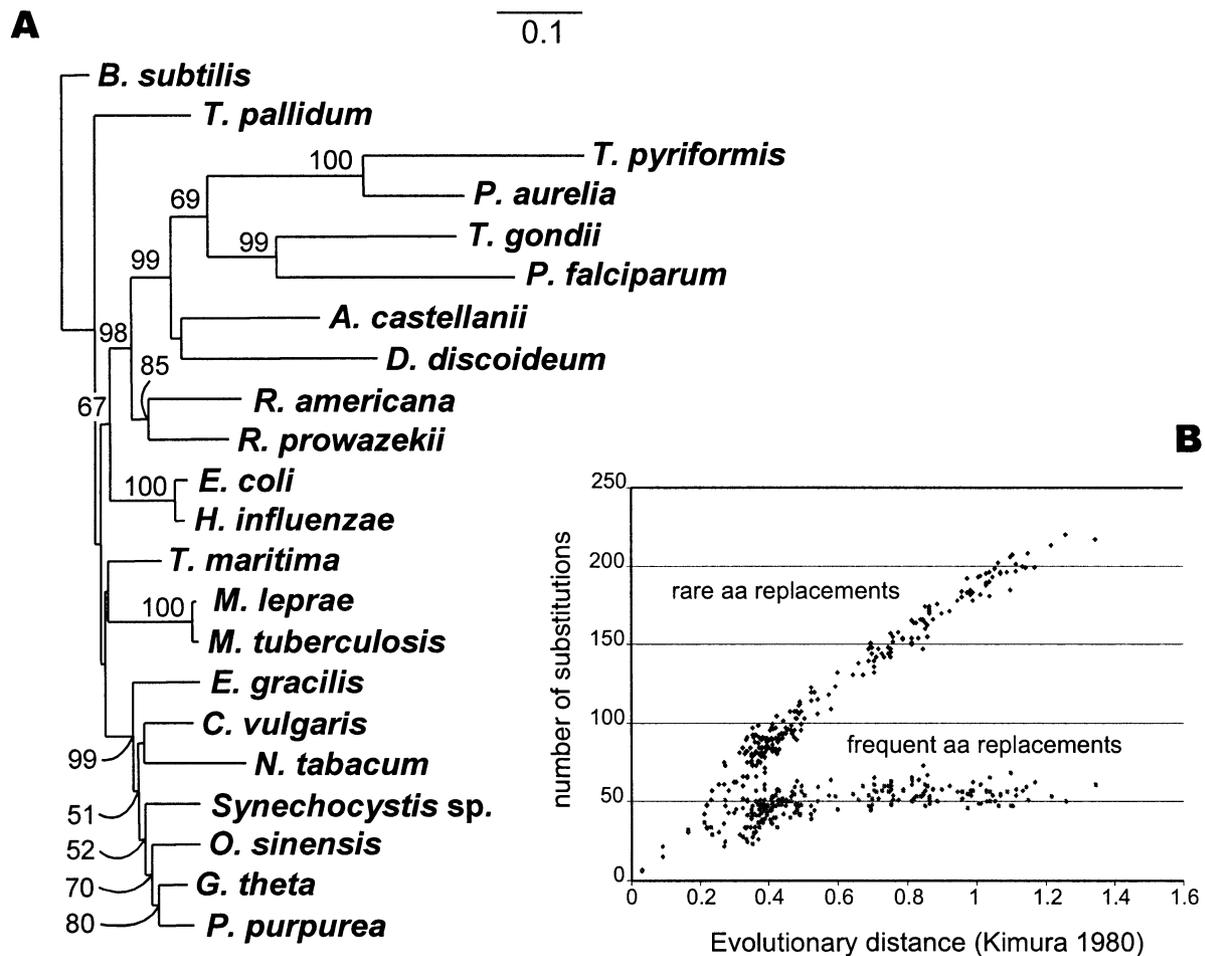


Fig. 2. (A) Distance neighbor-joining tree based on unsaturated aa positions of concatenated *rpl2*, *rpl14*, and *rps12* proteins. (B) Saturation of amino acids in *rpl2*, *rpl14*, and *rps12*.

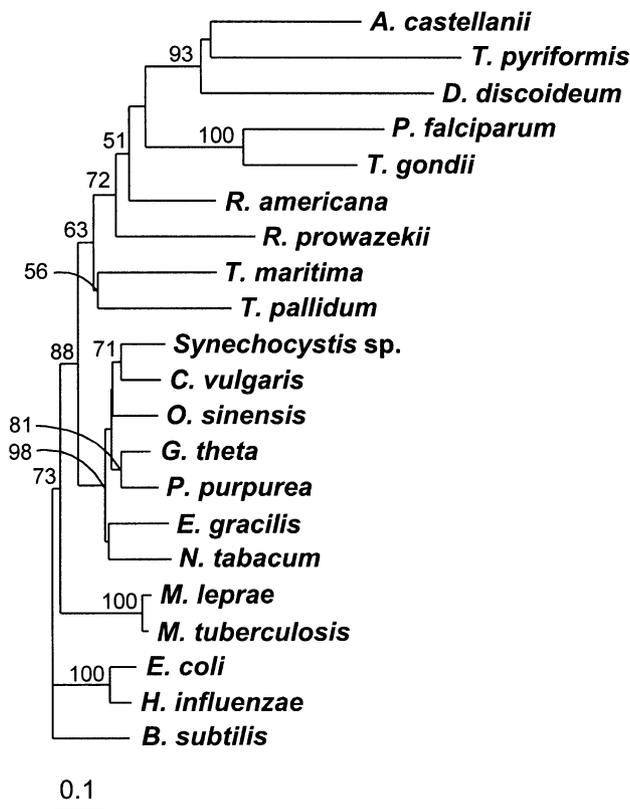


Fig. 3. Distance nt-based LogDet tree inferred from 560 variable nt characters (third codon position excluded) of concatenated *rpl2*, *rpl14*, and *rps12* sequences with the minimum evolution score of 4.66840. Bootstrap values (1000 replicates) are indicated.

plast *rps* within the mitochondrial cluster was confirmed by the Kishino-Hasegawa test (Table 3).

Even after the exclusion of the third codon position, the AT contents of the apicoplast *rp* sequences remain high, namely 74% for *Plasmodium* and 69% for *Toxoplasma*. These values differ considerably from the bacterial sequences, but are quite similar to the mitochondrial sequences, which are usually AT-rich. It has been shown that such a bias in sequence content can cause serious artifacts in the tree topology, since sequences with similar nucleotide composition tend to attract each other (Lockhart et al., 1992). Several methods have been devised to cope with this problem and, in particular, the LogDet method is supposed to circumvent the artifacts related with the variable base composition of sequences (Lockhart et al., 1994). In a tree computed on the basis of the LogDet distances, the apicoplast sequences also cluster with the mitochondrial ones (Fig. 3). Furthermore, even when all other mitochondrial sequences were omitted from the analysis, the apicoplast sequences clustered with *Reclinomonas* in 78% of the bootstrap samples. Similar results were obtained when evolutionary distances were computed according to Galtier and Gouy (1995). The latter estimation of evolutionary distances does not assume homogeneity or stationarity of the evolutionary process and should therefore also be useful

for phylogenetic analyses when compositional biases (transitions, transversions, GC bias) are observed in the data set.

A simpler method to account for bias in the GC or AT content is the so-called transversion analysis, where in the computation of evolutionary distances transitions are not taken into account (Swofford et al., 1996). It has been shown previously that the transversion analysis is less sensitive to the substitutional bias (Van de Peer et al., 1996). Again, a tree constructed on the basis of evolutionary distances computed according to Van de Peer et al. (1996) shows a tree topology similar to that shown in Fig. 3, with high bootstrap support for monophyletic grouping of apicoplasts and mitochondria.

#### 4. Discussion

##### 4.1. Extremely reduced apicoplast genome aggravates phylogenetic analyses

Apicomplexan parasites represent unicellular organisms holding really unique organelles. Both their plastid and mitochondrion contain the smallest organellar genomes known, extremely reduced during their evolution. When compared to  $\alpha$ -proteobacteria (1100 kb) and cyanobacteria (3500 kb) sharing ancestry with mitochondria and plastids, respectively, organellar genomes in apicomplexans represent only 0.5% (mitochondrion) and 1% (apicoplast) of their supposed ancestral genomes. Their uniqueness is further underlined by the presence of a single plastid and mitochondrion per apicomplexan cell and their permanent close proximity (Hopkins et al., 1999). Such a genome reduction is associated with a very rapid evolution of the apicoplast genes and anomalous nucleotide composition, especially in the case of the apicoplast ribosomal proteins, where the AT content rose over 85% and in some genes (e.g. *rpl11*) reached 97%. Nucleotide composition biased to such an extent complicates phylogenetic analysis and the danger of erroneous phylogeny is dramatically increasing.

On the other hand, the apicoplast sequences have so far been frequently used for phylogenetic analyses (e.g. Köhler et al., 1997; Blanchard and Hicks, 1999; Zhang et al., 2000). The most recent analysis used the apicoplast *rRNA* genes together with even more rapidly evolving genes of the dinoflagellate plastids (Zhang et al., 2000). Being aware of the long branch attraction phenomenon, the authors presented trees in which plastids and mitochondria formed highly supported distinct clades, which both contained rapidly evolving AT-rich sequences (Zhang et al., 2000). Nucleotide composition of the apicoplast *lsu rRNA* gene is comparable with that of *rp* genes used in our analysis (see Table 4). Other apicoplast genes, such as the *tufA* gene are, when compared to other plastids, AT-rich as well (see Table 4). This gene was retained in only one mitochondrial genome studied so far (*R. americana*), and in the tree it clustered together with the  $\alpha$ -proteobacterial homolog, while all plas-

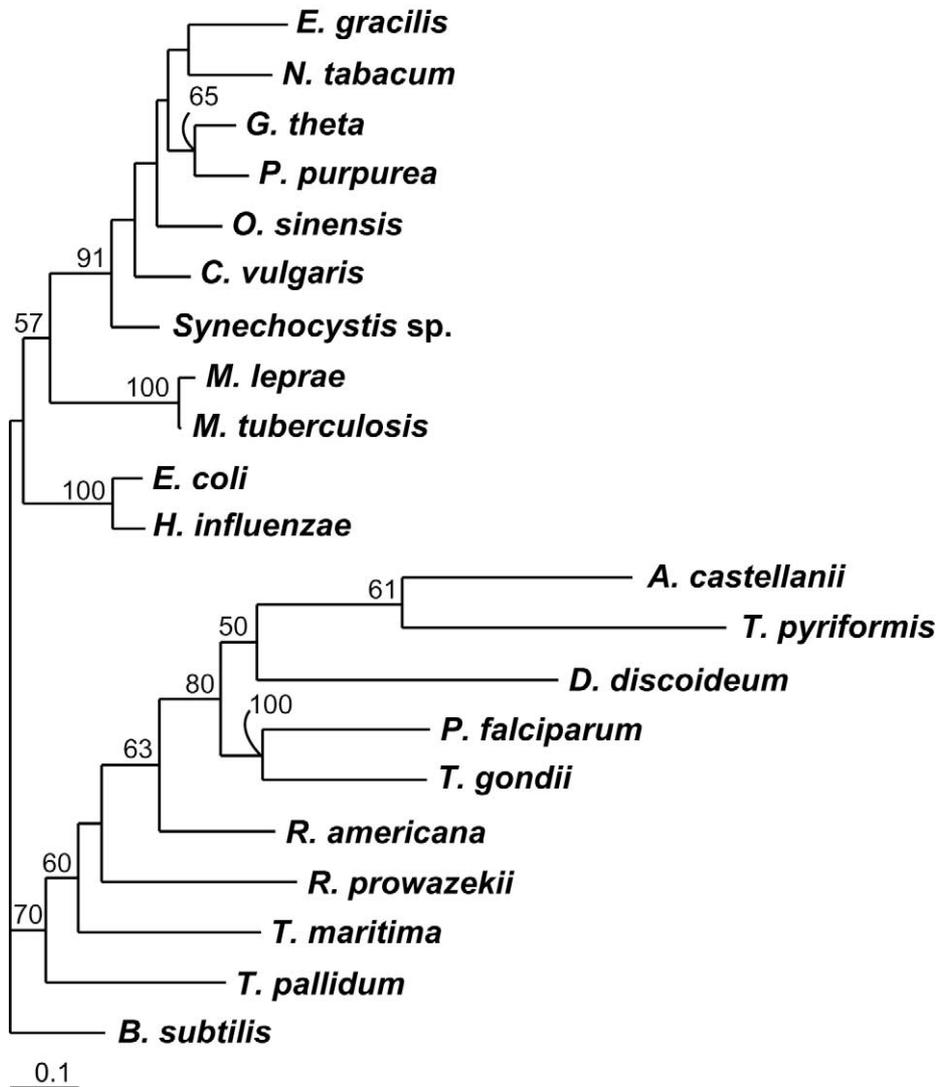


Fig. 4. Maximum likelihood nt-based tree inferred from 899 nt characters (third codon position excluded) of concatenated *rpl2*, *rpl14*, and *rps12* protein sequences. The best In score = 9704.61704.

tids including the apicoplasts formed a distinct clade (data not shown).

#### 4.2. Apicoplast *rpl2*, *rpl14* and *rps12* genes show a mitochondrial signal

To minimize the possibility of artifactual branching we

Table 3  
Kishino–Hasegawa test of hypothetical tree topologies<sup>a</sup>

Topology	$-\ln L$	Diff $-\ln L$	<i>P</i>
A	9825.78008	44.18237	0.008
B	9781.59771	Best	Best

<sup>a</sup> In the tree topology A the apicoplast sequences form common clade with plastids, while in the tree B they are affiliated with mitochondria. Kishino–Hasegawa test was calculated using RELL bootstrap (1000 replicates), two-tailed test with estimated ti/tv ratio and gamma shape parameter (PAUP\*).

have used the most conservative *rps* available. Some of the apicoplast *rp* genes differ from their plastid, mitochondrial and bacterial homologs to such an extent that the identification of these genes remains doubtful. Only some of the apicoplast *rp* genes are alignable, and from these, three proteins with homologs available in an adequately large set of organisms have been selected. Since these genes are relatively short, we have constructed concatenated alignments; however, the single-gene analyses led to the same, but weakly supported trees (Fig. 5). The analyses of the concatenated aa alignment show very consistent results, with both apicoplasts appearing within the mitochondrial cluster, close to the ciliate mitochondria. The position of the apicoplast *rp* genes within the mitochondrial cluster is quite stable and well supported. The transfer of analyzed *rp* genes from the mitochondrion into the plastid may explain such branching order. However, the ciliate mitochondrial sequences are highly divergent when compared to other

Table 4  
Nucleotide composition (AT content) of selected organellar and nuclear-encoded genes<sup>a</sup>

	<i>ssu rRNA</i>	<i>lsu rRNA</i>	<i>rpl2</i>	<i>rpl14</i>	<i>rps12</i>	<i>rps8*</i>	<i>rps9*</i>	<i>rpl28*</i>	<i>tufA</i>	Genome size (kb)
<i>P. falciparum</i> (pl)	76.97	80.30	83.47	87.90	81.27	94.06	69.21 <sup>N</sup>	?	78.83	35
<i>T. gondii</i> (pl)	64.01	69.00	78.23	83.33	78.18	43.49 <sup>N</sup>	46.98 <sup>N</sup>	47.46 <sup>N</sup>	72.55	35
<i>A. longa</i> (pl)	68.20	66.19	69.94	77.51	67.62	73.89	77.00	?	69.67	73
<i>E. gracilis</i> (pl)	58.77	57.73	65.35	65.57	62.96	71.39	68.40	?	64.40	143
<i>O. sinensis</i> (pl)	51.18	54.41	61.96	64.48	62.99	70.18	67.15	?	61.14	119
<i>P. purpurea</i> (pl)	48.36	51.73	62.03	63.11	62.93	61.94	63.77	69.76	62.72	191
<i>C. vulgaris</i> (pl)	47.31	50.41	57.00	60.16	59.68	65.23	64.73	?	61.06	151
<i>A. castellanii</i> (mt)	62.41	66.13	65.49	68.97	67.19	72.66	?	?	?	41
<i>D. discoideum</i> (mt)	62.35	65.59	72.91	73.91	72.34	78.21	?	?	?	55
<i>R. americana</i> (mt)	56.87	57.62	69.02	69.92	67.68	78.29	?	?	68.16	70
<i>T. pyriformis</i> (mt)	67.87	72.40	78.71	74.44	69.73	?	?	?	?	47

<sup>a</sup> Asterisks indicate *rp* genes with different location (<sup>N</sup>nuclear or organellar). Genes marked by “?” are probably located in the host nucleus.

mitochondria, and clustering of these sequences with apicoplasts may be an artifact. Although such artifactual phylogeny has to be taken into account, we conclude that all our phylogenetic analyses identified a mitochondrial signal in the apicoplast ribosomal protein genes *rpl2*, *rpl14* and *rps12*.

#### 4.3. Other non-photosynthetic plastids

The set of non-photosynthetic plastids known so far is very narrow and contains plastids from apicomplexans, the euglenoid *Astasia longa*, the heterotrophic unicellular algae *Prototheca wickerhamii* and *Polytoma uvella*, and the holoparasitic flowering plants (Gockel and Hachtel,

2000). The processes that follow after the loss of the photosynthetic function might be, in some aspects, very similar. The non-photosynthetic plastid genomes have been reduced in size; however, in the case of *A. longa* this reduction has been less extensive. *A. longa* represents together with its photosynthetically active relative *E. gracilis* an excellent model to study genome changes associated with the disruption of photosynthesis (their complete plastid sequences are already available) (Hallick et al., 1993; Gockel and Hachtel, 2000). The loss of photosynthesis caused a reduction of about 50% of the *A. longa* plastid genome (Table 4). Expectedly, all genes involved in photosynthesis, with the exception *rbcL*, have been eliminated from the *A. longa* plastid, and these genes are also absent from the apicoplast.

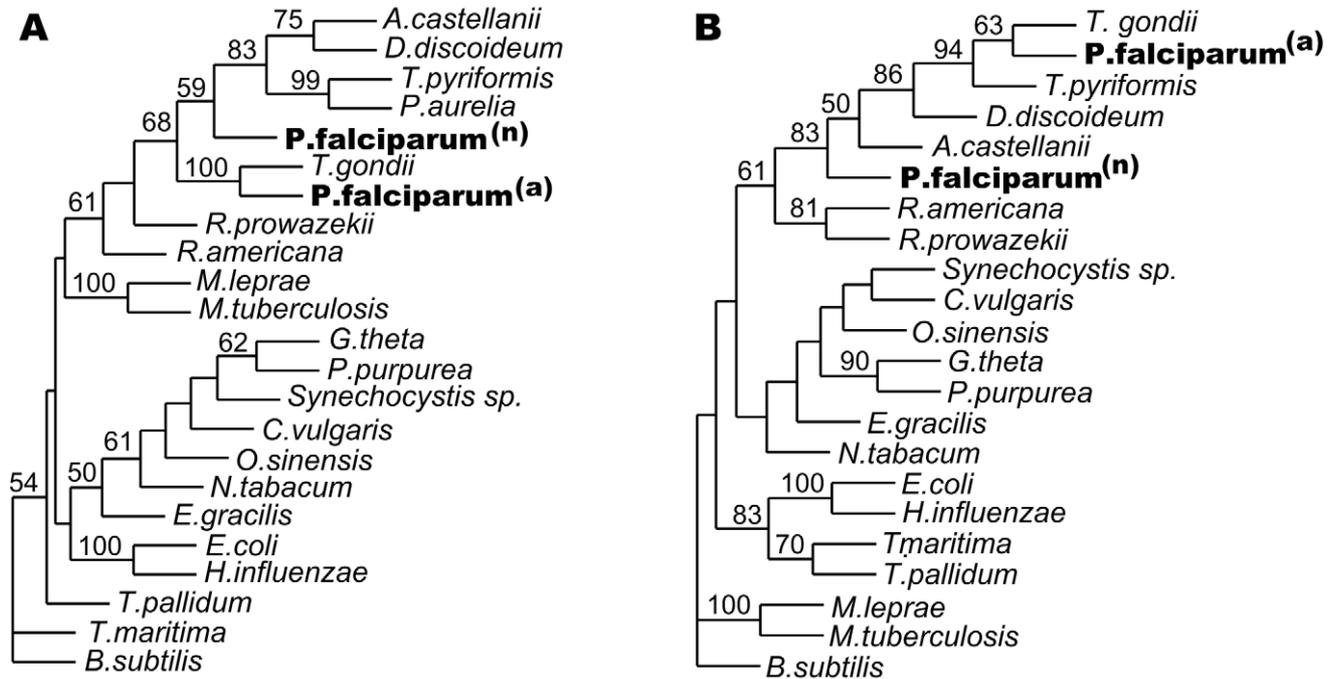


Fig. 5. Distance (minimum evolution, mean character differences with starting tree obtained via Neighbor-joining) single-gene based trees inferred from: (A) *rps12* aa sequences (ME score = 3.40054); (B) *rpl2* aa sequences (ME score = 4.70615). Both trees contain nuclear encoded (n) and apicoplast (a) *rp* genes. Bootstrap values (200 replicates) higher than 50% are indicated.

The nucleotide composition of the *A. longa* plastid genome has been significantly shifted in favor of AT bases (from 73.9% in *E. gracilis* to 77.6% AT for total genome; for *rp* genes see Table 4). Although genomes of all non-photosynthetic plastids may not behave in exactly the same way, features like their genome size, gene content and nucleotide composition may reflect the loss of photosynthesis. Although similar trends affected the evolution of apicoplast and the *A. longa* plastid, *rps* from the later genome appear within the plastid clade, while apicoplast *rps* clustered with mitochondria.

#### 4.4. Plastid origin of the apicoplast

Although the apicoplast genome seems to contain several possibly mitochondrially-derived genes, there is ample evidence that the organelle is a plastid. Its multimembranous ultrastructure, the absence of crista, the membrane lipid synthesis (Waller et al., 2000), the presence of the plastid isoprenoid biosynthesis pathway (Jomaa et al., 1999), and its existence within the cell together with a morphologically characteristic mitochondrion strongly favor the plastid origin of the apicoplast. Moreover, the *clpC* and *ycf24* genes are typical for plastids and are absent from all studied mitochondria. The apicoplast *tufA* gene, the mitochondrial homolog of which is present only in the primitive mitochondrion of *R. americana*, groups with plastids and cyanobacteria (Köhler et al., 1997). When aligned with mitochondrial and plastid homologs from organisms representing all relevant lineages, the apicoplast *ssu* and *lsu rRNA* genes branch, in accordance with previous studies (Blanchard and Hicks, 1999; Zhang et al., 2000), within the plastid lineage, although their position is unstable (data not shown). Despite extensive reduction, the gene order of the apicoplast *rps* in *Plasmodium* strongly resembles that found in the red algae, and was therefore inferred to be of the rhodophyte origin (McFadden et al., 1997).

#### 4.5. The apicoplast might contain a hybrid genome

The question, why do some apicoplast genes display mitochondrial origin, while others are apparently affiliated with plastids, remains to be answered. Our analyses suggest that individual apicoplast genes can be derived from both the mitochondrial and plastid genomes and apicoplast genome thus may constitute a hybrid assembly. Might such a genomic patchwork arise? The inter-organellar genetic exchange is reflected in several mitochondrial genomes which harbor mitochondrial, nuclear, and chloroplast genes, though only unidirectional transfer from the plastid into the mitochondrion of the *tRNA* genes has been documented (Marienfeld et al., 1999). Phylogenetic and circumstantial evidence support the possibility of direct gene transfer from the apicomplexan mitochondrion to the plastid for the *rp* genes. In the apicomplexan cell, both single-copy organelles are tightly adhered, and remain in close proximity for the entire life cycle (Hopkins et al.,

1999; Waller et al., 2000). The possibility of intracellular communication between these two organelles has inspired speculation on metabolic and other interactions (Hopkins et al., 1999). Due to the character of apicomplexan organelles, it can be speculated that these might share some components of their metabolic apparatus, such as ribosomal proteins. In such a case, disruption of apicoplast would influence the mitochondrion thus rendering the non-photosynthetic plastid essential for the host cell.

Candidates of mitochondrial homologs for *rps12* and *rpl2* are present in the nucleus of *P. falciparum*. The nuclear *rps12* gene contains two potential leader sequences, one of 140 aa at the N-terminus, and another of about 70 aa at the C-terminus. Our analyses show an affiliation of this nuclear-encoded gene with the apicoplast *rps12* within the mitochondrial cluster suggesting a possible gene duplication event (Fig. 5A). Both the nuclear-encoded *rpl2* gene that contains a typical mitochondrial leader and its plastid-encoded homolog are possibly of mitochondrial origin as well. However, both genes seem to have different mitochondrial ancestors (Fig. 5). Two mitochondrial genomes, the apicomplexan and algal, have been available to the apicomplexan cell during its evolution. Limited phylogenetic information contained in a single *rpl2* gene is not sufficient for the assignment of specific mitochondrial ancestor (Fig. 5). Our analyses of the concatenated aa alignment suggest that apicoplast *rp* genes may be derived from the apicomplexan mitochondrion (Fig. 1); however, the nt-based analysis does not support this relationship and places these genes close to the root of the mitochondrial cluster. To map relations between the organellar nuclear-encoded and apicoplast genes is not a trivial task. Sequences of these genes are in both cases influenced by their location. It has already been mentioned that the apicoplast genes are highly divergent and that their AT content has been accelerated. The nuclear-encoded genes of organellar origin are apparently affected by their new location (see Table 4). Therefore, if one copy of the duplicated mitochondrial gene is transferred to the apicoplast, and the other copy to the nucleus, the accumulated sequence differences would obscure their common origin. In this particular case, where two potential donors of the ancestral mitochondrial gene exist, a plethora of possible transient locations during their evolutionary history includes algal mitochondrion, algal nucleus, algal plastid, apicomplexan mitochondrion, apicomplexan nucleus, and apicoplast. Moreover, the evolutionary history of nuclear-encoded *rpl2* and *rps12* has likely been different.

Since the gene order of the apicoplast ribosomal operon strongly resembles that of the red alga (McFadden et al., 1997; Blanchard and Hicks, 1999) we suppose that the *rp* genes were transferred individually from mitochondrion to the plastid thus respecting the recipient operon structure. Such a scenario should be analogous to the transfer of organellar genes to the nucleus, where the nuclear-encoded organellar *rp* genes are probably transferred individually, because each such protein needs its own import sequence.

Individual transfer of *rp* genes from the organelle to the nucleus is also supported by different composition of the ribosomal operon in the apicoplasts, where *rps8* is still located in the plastid genome of *Plasmodium*, while in *Toxoplasma* it has already been transferred to the nucleus (Table 1).

While the independent monophyletic origin of plastids and mitochondria is well supported (Lang et al., 1997; Moreira et al., 2000), a possible inter-organelle gene transfer, as suggested here for the apicomplexan mitochondrion and plastid, should be taken into consideration when reconstructing the evolutionary histories of both organelles.

## Acknowledgements

This work was supported by Grant Agency of the Czech Academy of Sciences, grants no. B5022904 and A6022903, and by the Ministry of Education of the Czech Republic, project CEZ:AJ06/98:F1-123100003. Y.V.d.P. is a Research Fellow of the National Fund for Scientific Research (Belgium) and acknowledges financial support from the University of Konstanz (Germany) and the University of Antwerp (Belgium). The authors thank Pete Lockhart for the LogDet analyses of the concatenated aa sequence alignment, and Jiří Vávra and Lys Guilbride for helpful discussions and critical reading of the manuscript.

## References

- Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., Doolittle, W.F., 2000. A kingdom level phylogeny of eukaryotes based on combined protein data. *Science* 290, 972–977.
- Blanchard, J.L., Hicks, J.S., 1999. The non-photosynthetic plastid in malarial parasites and other apicomplexans is derived from outside the green plastid lineage. *J. Eukaryot. Microbiol.* 46, 367–375.
- Cavalier-Smith, T., 1993. Kingdom *Protozoa* and its 18 phyla. *Microbiol. Rev.* 47, 953–994.
- Galtier, M., Gouy, M., 1995. Inferring phylogenies from DNA sequences of unequal base composition. *Proc. Natl. Acad. Sci. USA* 92, 11317–11321.
- Gockel, G., Hachtel, W., 2000. Complete gene map of the plastid genome of the nonphotosynthetic euglenoid flagellate *Astasia longa*. *Protist* 151, 347–351.
- Hallick, R.B., Hong, L., Drager, R.G., Favreau, M.R., Monfort, A., Orsat, B., Spielmann, A., Stutz, E., 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.* 21, 3537–3544.
- Hepler, P.K., Huff, C.G., Sprinz, H., 1966. The fine structure of the exoerythrocytic stages of *Plasmodium fallax*. *J. Cell Biol.* 30, 333–358.
- Hopkins, J., Fowler, R., Krishna, S., Wilson, I., Mitchell, G., Bannister, L., 1999. The plastid in *Plasmodium falciparum* asexual blood stages: a three-dimensional ultrastructural analysis. *Protist* 150, 283–295.
- Jomaa, H., Weisner, J., Sanderbrand, S., Altincicek, B., Weidenmeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H.K., Soldati, D., Beck, E., 1999. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* 285, 1573–1575.
- Kilejian, A., 1975. Circular mitochondrial DNA from the avian malarial parasite *Plasmodium lophurae*. *Biochim. Biophys. Acta* 390, 276–284.
- Köhler, S., Delwiche, C.F., Denny, P.W., Tilney, L.G., Webster, P., Wilson, R.J.M., Palmer, J.D., Roos, D.S., 1997. A plastid of probable green algal origin in apicomplexan parasites. *Science* 275, 1485–1488.
- Lang, B.F., Burger, G., Okelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., Gray, M.W., 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* 387, 493–497.
- Lockhart, P.J., Penny, D., Hendy, M.D., Howe, C.J., Beanland, T.J., Larkum, A.W., 1992. Controversy on chloroplast origins. *FEBS Lett.* 301, 127–131.
- Lockhart, P.J., Steel, M.A., Hendy, M.D., Penny, D., 1994. Recovering evolutionary trees under more realistic model of sequence evolution. *Mol. Biol. Evol.* 11, 605–612.
- Marienfeld, J., Unseld, M., Brennicke, A., 1999. The mitochondrial genome of *Arabidopsis* is composed of both native and immigrant information. *Trends Plant Sci.* 4, 495.
- McFadden, G.I., Waller, R.F., Reith, E., Lang-Unnasch, N., 1997. Plastids in apicomplexan parasites. *Plant Syst. Evol.* 11, 261–287.
- Moreira, D., Le Guyader, H., Philippe, H., 2000. The origin of red algae and the evolution of chloroplasts. *Nature* 405, 69–72.
- Penny, D., Hasegawa, M., Waddell, P.J., Hendy, M.D., 1999. Mammalian evolution: timing and implications from using the LogDeterminant transform for proteins of differing amino acid composition. *Syst. Biol.* 48, 76–93.
- Schrével, J., 1971. Observations biologiques et ultrastructurales sur les Selenidiidae et leurs conséquences sur la systématique des Grégariomorphes. *J. Protozool.* 18, 448–470.
- Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13, 964–969.
- Swofford, D.L., 1998. PAUP\* Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4, Sinauer Associates, Sunderland, MA.
- Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.). *Molecular Systematics*, Sinauer Associates, Sunderland, MA.
- Vaidya, A.B., Akella, R., Suplick, K., 1989. Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase pair DNA of a malarial parasite. *Mol. Biochem. Parasitol.* 35, 97–108.
- Van de Peer, Y., De Wachter, Y., 1994. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput. Appl. Biosci.* 10, 569–570.
- Van de Peer, Y., Rensing, S., Maier, U.G., De Wachter, R., 1996. Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proc. Natl. Acad. Sci. USA* 93, 7732–7736.
- Van de Peer, Y., Baldauf, S., Doolittle, W.F., Meyer, A.A., 2000. A comprehensive phylogeny of the rRNA crown eukaryotes based on rate calibrated evolutionary distances. *J. Mol. Evol.* 51, 565–576.
- Van de Peer, Y., Frickey, T., Taylor, J.S., Meyer, A., 2002. Dealing with saturation at the amino acid level: A case study based on anciently duplicated zebrafish genes. *Gene* (in press).
- Waller, R.F., Reed, M.B., Cowman, A.F., McFadden, G.I., 2000. Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *EMBO J.* 19, 1794–1802.
- Zhang, Z.D., Green, B.R., Cavalier-Smith, T., 2000. Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: A possible common origin for sporozoan and dinoflagellate plastids. *J. Mol. Evol.* 51, 26–40.