Supplemental data for:

PNAS | **August 28, 2001** | vol. 98 | no. 18 | **10202-10207**

Molecular phylogenetic analyses of the mitochondrial ADP-ATP carriers: The Plantae/Fungi/Metazoa trichotomy revisited

Ari Löytynoja and Michel C. Milinkovitch*

Unit of Evolutionary Genetics, Free University of Brussels (ULB), C.P. 300, Institute of Molecular Biology and Medicine, Rue Jeener and Brachet 12, B-6041 Gosselies, Belgium
http://www.ulb.ac.be/sciences/ueg

Supplemental Text

The substrates transported by mitochondrial carrier family (MCF) members are involved in functions and processes as crucial as β oxidation of AcylCoA, tricarboxylic acid cycle, and oxidative phosphorylation. Structurally related but functionally somewhat different members of the MCF include: the heat-producing uncoupling protein exclusively found in the mammalian brown fat tissue and the mtRNA splicing protein proteins involved in RNA splicing (1, 2).

Materials and Methods

Data. The ADP–ATP carrier (AAC) sequences were searched in nonredundant DNA GenBank with the program TBLASTN with the query sequences AB008463 (*Rana rugosa*), M57424 (*Homo sapiens*), Y10618 (*Drosophila melanogaster*), X76112 (*Caenorhabditis elegans*), D83069 (*Halocynthia roretzi*), M76669 (*Chlorella kessleri*), AF100676 (*Dictyostelium discoideum*), U04335 (*Plasmodium falciparum*), M12514 (*Saccharomyces cerevisiae*), X00363 (*Neurospora crassa*), X65549 (*Arabidopsis thaliana*), AL021749 (*A. thaliana*), X65194 (*Chlamydomonas reinhardtii*), and AF049130 (*Trypanosoma brucei*). For cDNA sequences, the flanking noncoding regions were removed and the remaining nucleotide sequence translated. For nuclear DNA, the predicted introns were spliced out and the resulting DNA sequence translated. In some cases, because sequence annotation was insufficient, we predicted the coding region and its ORF with the program WISE2 (Ewan Birney, Sanger Centre, Cambridge, U.K.) by using one of the query peptides as a reference. During removal of redundant sequences, WISE2-predicted ORFs were preferentially removed in redundant pairs, because the WISE2 algorithm has a tendency to misinterpret the edges of the coding regions. Two aberrant sequences (*A. thaliana*, AB009049; *C. kessleri,* M76669) were removed after a brief preliminary study. Indeed, AB009049 did not group with other plant or *Arabidopsis* genes but tended to position itself at the bottom of the ingroup. Additionally, although it had not been isolated from cDNA libraries, the original sequence has no intron and contains an adenine-rich region some 300 nucleotides downstream from the STOP codon. We could identify a full ORF, but no corresponding expressed sequence tag was found in GenBank. The sequence is likely to correspond to an event of retrotransposition. Similarly, M76669 from *C. kessleri* is clearly an AAC member, but it is abnormally divergent from any plant, even from the closely related green alga (protophyte) *C. reinhardtii*. In preliminary phylogenetic trees,

M76669 tended to locate itself among the deepest nodes of the ingroup together with *Plasmodium*. The reason for this large divergence remains unclear, but a sequencing artifact or a horizontal gene transfer might be expected. Reinclusion of AB009049 and M76669 had no significant impact on the phylogenetic results discussed in the manuscript.

Alignment. The plant AAC sequences include a long trailing sequence not found in other eukaryotes. This fragment is therefore uninformative for the question at hand, and the corresponding 18 stable but noninformative sites were removed from the multiple alignments. All alignments used in these analyses are available at http://dbm.ulb.ac.be/ueg.

Comparison with 18S. From the original extensive 18S data set (3), we selected a taxon sampling manageable for maximum likelihood (ML) analyses and rooted the ingroup with three archaean sequences (making 30 taxa, 3,276 bp). However, alignment of ribosomal RNA sequences is notoriously difficult. Our soap sensitivity analysis indeed detected a large proportion of positions unstable to variations of the alignment parameters. Hence, the alignment was reduced to 761 nucleotides after computing the strict consensus among the original alignment from ref. 3 and five different alignments generated with soap [(4), opening/extension penalties of 14/6, 15/5, 15/6, 15/7, 16/6]. As the 18S data consist of nucleotide sequences, a heuristic ML search was conducted with PAUP* (5) rather than protml. Analytical parameters were as follows: four substitution rate categories, estimating base frequencies, proportion of invariable sites, γ–shape parameter, and substitution rate matrix (general time-reversible model) from a neighbor-joining (NJ) tree produced with Tamura–Nei distances. The search was performed with tree bisection and reconnection (TBR) branch swapping starting from the NJ tree. Thirty searches with TBR swapping starting from random (rather than NJ) trees failed to find better trees. Signal decay analyses were also performed.

Results and Discussion

Phylogenetic Analyses. The figures illustrating (*i*) signal decay analysis of the AAC data set (i.e., sequential removal of ML γ–rate categories of sites; see *Materials and Methods*), and (*ii*) reanalysis of an 18S data set before and after removal of sites unstable to variations of the alignment parameters are provided as Figs. 4 and 5, respectively.

1. Nelson, D. R., Felix, C. M. & Swanson, J. M. (1998) *J. Mol. Biol.* **277,** 285–308.

2. Walker, J. E. (1992) *Curr. Opin. Struct. Biol.* **2,** 519–526.

3. Van de Peer, Y. & De Wachter, R. (1997) *J. Mol. Evol.* **45,** 619–630.

4. Löytynoja, A. & Milinkovitch, M. C. (2001) *Bioinformatics* **17,** 573–574.

5. Swofford, D. L. (2000) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods) (Sinauer, Sunderland, MA), Ver. 4.

Fig. 4. Best maximum likelihood trees (exhaustive search with same constraints as in Fig. 2) obtained through analyses of the AAC mitochondrial carrier data set after sequentially removing 0–6 rate classes from the alignment. The [Plantae + Fungi] clade is stable to this sequential exclusion of the six fastest evolving partitions of sites. The efficiency of this procedure for reducing noise is exemplified by *Plasmodium* and *Trypanosoma* (indicated by arrows), which are taking their expected phylogenetic positions when the fastest category of sites is removed. Scale bars indicate 10 substitutions per site.

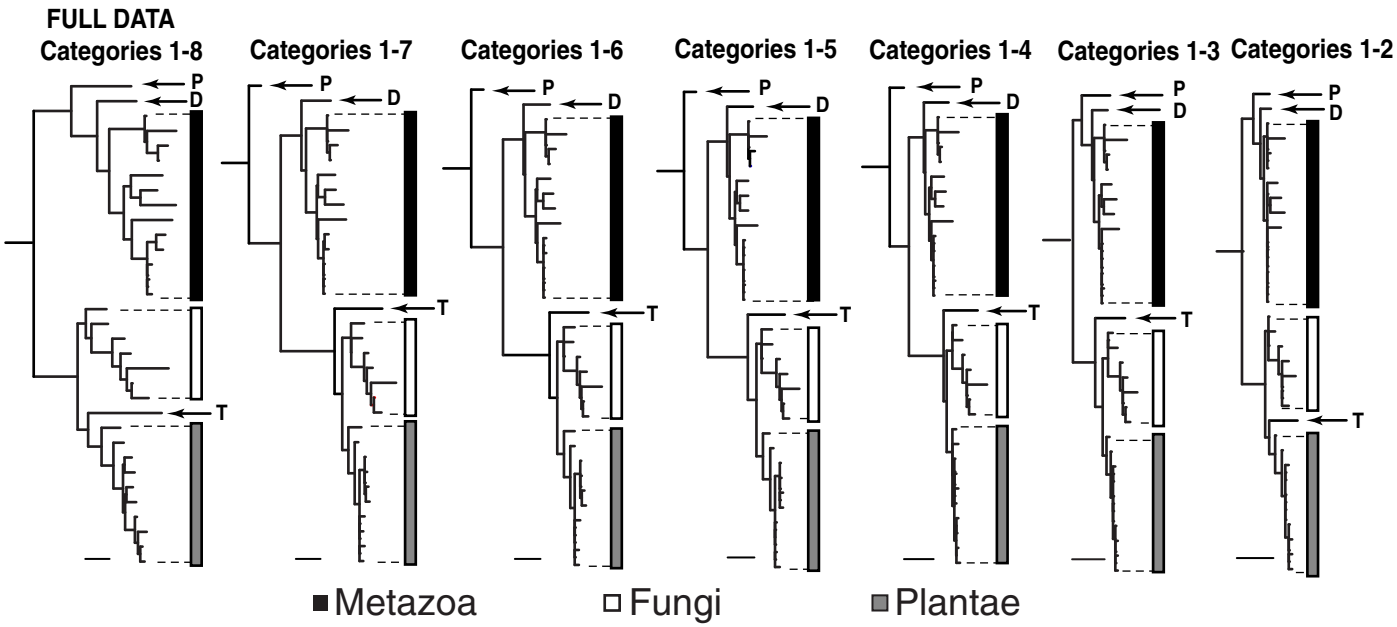
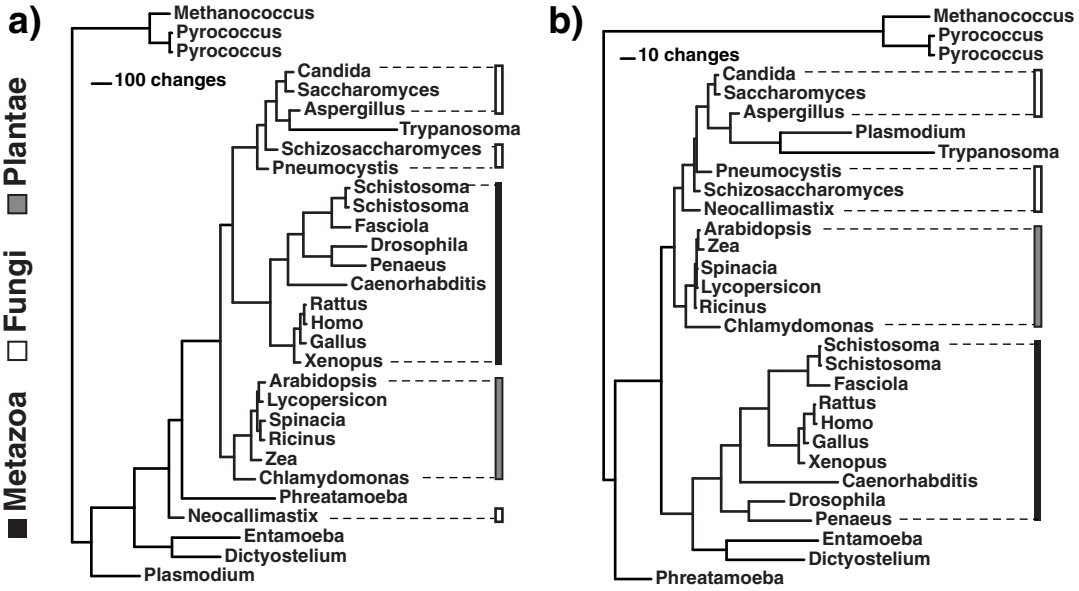

FULL DATA
Categories 1-8
Categories 1-7
Categories 1-6
Categories 1-5
Categories 1-4
Categories 1-3
Categories 1-2
P
D
T
Metazoa
Fungi
Plantae

Fig. 5. Heuristic maximum likelihood analyses of the 18S nucleotide data set. (*a*) Analysis of the original alignment yields a monophyletic [Metazoa + Fungi] invaded by *Trypanosoma*. (*b*) Analysis of the same alignment after removal of positions unstable to alignment parameters (by using SOAP; see *Materials and Methods*) yields a [Plantae + Fungi + *Trypanosoma*] clade invaded by *Plasmodium*.


a)
Plantae
Fungi
Metazoa
Methanococcus
Pyrococcus
Pyrococcus
100 changes
Candida
Saccharomyces
Aspergillus
Trypanosoma
Schizosaccharomyces
Pneumocystis
Schistosoma
Schistosoma
Fasciola
Drosophila
Penaeus
Caenorhabditis
Rattus
Homo
Gallus
Xenopus
Arabidopsis
Lycopersicon
Spinacia
Ricinus
Zea
Chlamydomonas
Phreatamoeba
Neocallimastix
Entamoeba
Dictyostelium
Plasmodium
b)
Methanococcus
Pyrococcus
Pyrococcus
10 changes
Candida
Saccharomyces
Aspergillus
Plasmodium
Trypanosoma
Pneumocystis
Schizosaccharomyces
Neocallimastix
Arabidopsis
Zea
Spinacia
Lycopersicon
Ricinus
Chlamydomonas
Schistosoma
Schistosoma
Fasciola
Rattus
Homo
Gallus
Xenopus
Caenorhabditis
Drosophila
Penaeus
Entamoeba
Dictyostelium
Phreatamoeba