Phylogenetic Analysis of Mitochondrial Outer Membrane β-Barrel Channels

Małgorzata Wojtkowska[1,†], Marcin Jąkalski[2,†], Joanna R. Pieńkowska[3], Olgierd Stobienia[1], Andonis Karachitos[1], Teresa M. Przytycka[4], January Weiner 3rd[2], Hanna Kmita[1,*], and Wojciech Makałowski[2,*]

[1]Laboratory of Bioenergetics, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland

[2]Institute of Bioinformatics, Faculty of Medicine, University of Muenster, Germany

[3]Department of Cell Biology, Institute of Experimental Biology, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland

[4]NCBI, NLM, NIH Computational Biology Branch, Bethesda, Maryland

†These authors contributed equally to this work.

*Corresponding author: E-mail: kmita@amu.edu.pl; wojmak@uni-muenster.de.

Accepted: 2 December 2011

Data deposition: *Acanthamoeba castellanii* mRNA sequence of TOM40 was deposited in the GenBank under accession number HQ259301.

Abstract

Transport of molecules across mitochondrial outer membrane is pivotal for a proper function of mitochondria. The transport pathways across the membrane are formed by ion channels that participate in metabolite exchange between mitochondria and cytoplasm (voltage-dependent anion-selective channel, VDAC) as well as in import of proteins encoded by nuclear genes (Tom40 and Sam50/Tob55). VDAC, Tom40, and Sam50/Tob55 are present in all eukaryotic organisms, encoded in the nuclear genome, and have β-barrel topology. We have compiled data sets of these protein sequences and studied their phylogenetic relationships with a special focus on the position of Amoebozoa. Additionally, we identified these protein-coding genes in *Acanthamoeba castellanii* and *Dictyostelium discoideum* to complement our data set and verify the phylogenetic position of these model organisms. Our analysis show that mitochondrial β-barrel channels from Archaeplastida (plants) and Opisthokonta (animals and fungi) experienced many duplication events that resulted in multiple paralogous isoforms and form well-defined monophyletic clades that match the current model of eukaryotic evolution. However, in representatives of Amoebozoa, Chromalveolata, and Excavata (former Protista), they do not form clearly distinguishable clades, although they locate basally to the plant and algae branches. In most cases, they do not posses paralogs and their sequences appear to have evolved quickly or degenerated. Consequently, the obtained phylogenies of mitochondrial outer membrane β-channels do not entirely reflect the recent eukaryotic classification system involving the six supergroups: Chromalveolata, Excavata, Archaeplastida, Rhizaria, Amoebozoa, and Opisthokonta.

Key words: mitochondrial β-barrels, mitochondrial outer membrane, VDAC, Sam50/Tob55, Tom40, *Acanthamoeba castellanii*, *Dictyostelium discoideum*.

Introduction

The biogenesis and proper function of mitochondria require an import of 90–99% of mitochondrial proteins as well as a metabolite exchange with cytoplasm. These transport processes are mediated by proteins located in both mitochondrial membranes but are initiated by proteins from the mitochondrial outer membrane that display channel activity, namely voltage-dependent anion-selective channel (VDAC), Tom40, and Sam50/Tob55. VDAC, known also as mitochondrial porin, supports the flux of metabolites (Blachly-Dyson and Forte 2001; Colombini 2004; Shoshan-Barmatz et al. 2006, 2010; Mannella and Kinnally 2008; Rostovtseva and Bezrukov 2008). Although the functional diffusion channel is formed by a single VDAC protein, it has also been proven that VDAC proteins are able to form oligomeric complexes (Hoogenboom et al. 2007; Gonçalves et al. 2007). Tom40 and Sam50/Tob55 are crucial components of protein import complexes, namely the TOM complex (translocase of

© The Author(s) 2011. Published by Oxford University Press on behalf of the *Society for Molecular Biology and Evolution.*
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

the mitochondrial outer membrane) and the SAM/TOB complex (sorting and assembly machinery/topogenesis of the mitochondrial outer membrane β-barrel proteins), respectively (Dolezal et al. 2006; Bohnert et al. 2007; Neupert and Herrmann 2007; Becker et al. 2008; Walther et al. 2009). The SAM/TOB complex is essential for insertion of β-barrel integral proteins into the mitochondrial outer membrane as well as for biogenesis of the TOM complex. The TOM complex is regarded as a general entry gate for mitochondria and as the complex is responsible for decoding of targeting signals, translocation of imported proteins across or into the outer membrane, and their subsequent sorting.

Interestingly, VDAC, Tom40, and Sam50/Tob55 are predicted to have a β-barrel topology (Schulz 2000; Bay and Court 2002; Gentle et al. 2004; Ryan 2004; Rapaport 2005; Paschen et al. 2005; Zeth 2010). However, it should be noted that, to date, only the structure of mammalian VDAC1 has been solved and the remaining members have been assigned to the same family as VDAC only on the basis of both secondary structure content and in silico predictions (Walther et al. 2009; Zeth 2010). Moreover, although for mammalian VDAC, the structure was confirmed by nuclear magnetic resonance spectroscopy and X-ray crystallography (Bayrhuber et al. 2008; Hiller et al. 2008; Ujwal et al. 2008), its details are still under discussion (Colombini 2004). The β-barrel proteins are present in the outer membranes of Gram-negative bacteria and in the organelles of endosymbiotic origin, so their presence in the organellar outer membrane reflects the evolutionary origin of mitochondria and chloroplasts from endosymbiotic bacteria (Gray et al. 1999; Lister et al. 2005; Cavalier-Smith 2006; Dolezal et al. 2006; Perry et al. 2008; Kutik et al. 2009; Walther et al. 2009). However, the data concerning molecular phylogeny of VDAC, Tom40, and Sam50/Tob55, as well as number of genes coding for these proteins identified in different eukaryotic organisms, is rather limited. Current data allow for the following conclusions. First, amino acid sequence similarity to bacterial β-barrel proteins is observed only for Sam50/Tob55, what seems to confirm the origin of Sam50/Tob55 from the protein present in the original endosymbiont (Paschen et al. 2003, 2005; Gentle et al. 2004; Dolezal et al. 2006) and implies higher levels of amino acid sequence conservation when compared with Tom40 and VDAC. The latter two proteins probably evolved from the endosymbiont outer membrane proteins, but they appear to diverge too far in the amino acid sequence (Maćasev et al. 2004; Paschen et al. 2005; Cavalier-Smith 2006). Second, Tom40, together with Tom22 and Tom7, is regarded as the earliest component of the TOM complex (Maćasev et al. 2004; Lister et al. 2005; Dolezal et al. 2006; Perry et al. 2008). The complex, or at least its crucial subunits, seems to have derived early in the evolution of mitochondria, as it was probably essential for the establishment of the protein translocases of the mitochondrial inner membrane (Cavalier-Smith 2006; Dolezal et al. 2006). Third, VDAC gene was duplicated independently in different lineages of eukaryotic organisms several times during the evolution (Sampson et al. 1996; Saccone et al. 2003; Young et al. 2007).

It should be emphasized that phylogenetic data concerning VDAC, Tom40, and Sam50/Tob55 include mainly proteins identified for animals, fungi, and plants, and only few for unicellular eukaryotes commonly referred to as protists—a diverse group of eukaryotic microorganisms defined by exclusion of animals, fungi, and plants. Taking into account the polyphyletic character of protists, the historic division of eukaryotic organisms into four kingdoms of Plantae, Animalia, Fungi, and Protista has been replaced recently by a new system that consists of six large supergroups—Chromalveolata, Excavata, Archaeplastida, Rhizaria, Amoebozoa, and Opisthokonta (Adl et al. 2005; Keeling et al. 2005). Nevertheless, additional data are still required to support the six supergroups classification system before a common acceptance (Cavalier-Smith 2006; Adl et al. 2005; Keeling et al. 2005; Parfrey et al. 2006; Dacks et al. 2008).

This study focuses on the amoeba *Acanthamoeba castellanii*'s and the slime mold *Dictyostelium discoideum*'s β-barrel proteins that form channels in the mitochondrial outer membrane. Both microorganisms are currently proposed to be members of Amoebozoa (Dacks et al. 2002; Keeling et al. 2005; Russell et al. 2006; Shutt and Gray 2006) and are commonly applied as model organisms in different biological studies, including plant and animal mitochondria research (Jarmuszkiewicz et al. 2002; Kicinska et al. 2007), symbiont and predator–prey relationships, horizontal gene transfer (Anderson et al. 2005; Alsam et al. 2006; Watkins and Gray 2006), cell divisions, cell motility, phagocytosis, chemotaxis, signaling, and development (Eichinger and Noegel 2003; Williams et al. 2005). However, till now only Tom40 and VDAC of *D. discoideum* have been identified (Troll et al. 1992; Maćasev et al. 2004). The aim of this study was to identify Sam50/Tob55 of *D. discoideum* as well as VDAC, Tom40, and Sam50/Tob55 of *A. castellanii* and use them to perform comprehensive analysis of the evolutionary history of mitochondrial outer membrane channels of a predicted β-barrel topology. Given that amoebozoans are a sister group of fungi and Metazoa, which diverged from the animal/fungal line after its split from plants (Eichinger et al. 2005), they present common features to both lineages, alike in their mitochondrial physiology as well as protein repertoire. For these reasons, results of our phylogenetic studies might aid the verification of the position of *A. castellanii*, *D. discoideum*, and other amoebozoans on the eukaryotic tree of life. In other words, the second goal of this work was to check if mitochondrial outer membrane β-barrel channels, present in all eukaryotes, conserved in their structure and performed function, might serve as good phylogenetic markers providing a solution to eukaryotic organism classification.

Materials and Methods

Cell Culture and Isolation of the Outer Membrane Vesicles of *A. castellanii* Mitochondria

The amoeba *A. castellanii* (strain Neff) was cultured axenically as described by Jarmuszkiewicz et al. (1997). Trophozoites of the amoeba were collected at an early stationary phase (72 h after inoculation) and mitochondria were isolated according to the published procedure (Jarmuszkiewicz et al. 1997). The outer membrane vesicles (OMVs) were isolated and purified as described by Mayer et al. (1993).

Mass Spectrometry Analysis and Data Processing

The OMV of *A. castellanii* (50 μg) were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and analyzed by nanospray liquid chromatography tandem mass spectrometry (LC-MS/MS) (Zentrallabor für Proteinanalytik, Ludwig-Maximilians-Universität, München, Germany). For the identification by mass spectrometry, proteins of molecular weight (MW) close to that of VDAC, Tom40, and Sam50/Tob55 were excised from the SDS-PAGE gels. The obtained results were used to search for data derived from publicly available *A. castellanii* expressed sequence tags (ESTs), generated by the Protist EST Program (TBestDB, http://tbestdb.bcm.umontral.ca, O'Brien et al. 2007) and *A. castellanii*'s ESTs from the Baylor College of Medicine Human Genome Sequencing Center (http://www.hgsc.bcm.tmc.edu/microbial-detail.xsp?project_id=163 [last accessed 2011 Dec 18]). Peptide mass fingerprinting was achieved by Prospector (MS-digest) (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest [last accessed 2011 Dec 18]). Nucleotide sequence extension of the studied genes was obtained by collecting traces of sequences from trace archives against two databases: *A. castellanii* Whole Genome ShotGun and "*A. castellanii* other" (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PAGE=Nucleotides&PROGRAM=blastn [last accessed 2011 Dec 18]). Cap3 program (http://pbil.univ-lyon1.fr/cap3.php [last accessed 2011 Dec 18]) was used for sequence assembly. Amino acid sequences were obtained by basic Blast tools (e.g., BlastX) or GenScan (http://genes.mit.edu/GENSCAN.html [last accessed 2011 Dec 18]). Transmembrane β-strands in the analyzed sequences were identified by TBBpred (http://www.imtech.res.in/raghava/tbbpred/ [last accessed 2011 Dec 18], Natt et al. 2004) and TMBETA-NET (http://psfs.cbrc.jp/tmbeta-net/ [last accessed 2011 Dec 18], Gromiha et al. 2005). The prediction of protein domains was made using the Pfam search tool (http://pfam.sanger.ac.uk/search [last accessed 2011 Dec 18], Finn et al. 2010).

Polymerase Chain Reaction Analysis of *A. castellanii* TOM40

The 3′ end of cDNA coding for the Tom40 protein was studied by reverse transcription–polymerase chain reaction (RT-PCR) and rapid amplification of cDNA 3′ end by PCR (3′RACE) (Mullis et al. 1994) with gene-specific primers (F6tom 5′GGCAAATTCGATGATCTCTC3′ and F8tom 5′GATGTCTCACGCACCATCAC 3′) based on a partial nucleotide sequence of Tom40. To obtain the 3′ end, 1 μg of DNA-free total RNA was reverse transcribed (SuperScript II Reverse Transcriptase, Invitrogen) using 1 μl Q_T primer (10 mM) that annealed to the polyadenylate tail of mRNA. Next, 1 μl of the reverse transcription reaction was amplified using the forward primer F6tom and reverse primer Q_O that annealed to the 3′ end of Q_T. The first round of 3′RACE was conducted under the following cycling conditions: 95 °C for 3 min, 48 °C for 2 min, 72 °C for 40 min. The next 30 cycles were performed at 94 °C for 30 s, 55 °C for 30 s, 72 °C for 2 min, and at the end an incubation at 72 °C for 10 min was applied. An aliquot of the final product of 1,100 bp (0.5 μl) was taken to the second round of 3′RACE using "nested" primers—Q_I and F8tom under the following cycling conditions: 95 °C for 3 min and the next 30 cycles at 94 °C for 30 s, 58 °C for 30 s, 72 °C for 2 min, and at the end at 72 °C for 10 min. Template-free samples were used as negative controls in both PCR rounds. A product with the size of 1,000 bp was obtained and then purified and sequenced on a 3130xl Genetic Analyzer (Applied Biosystems Applera).

5′ end sequence of *TOM40* was confirmed by standard PCR using the gene-specific primers: forward F1tom (5′ ATGGACAAGCTTCCCTCACC 3′) and reverse RBtom (5′ GACTGGAGGAGGTTGTAGGC 3′). The thermocycling program consisted of one hold at 94 °C for 3 min, followed by 35 cycles for 30 s at 94 °C, 35 s at 60 °C, 1 min at 72 °C, and at 72 °C for 10 min at the end. Template-free samples were used as negative controls. The obtained 934 bp product was purified and then sequenced using a 3130xl Genetic Analyzer (Applied Biosystems Applera).

Identification of Tom40 Contigs

The *TOM40* nucleotide sequence (http://tbestdb.bcm.umontreal.ca/ [last accessed 2011 Dec 18], *A. castellanii* Cluster ACL00002106) was used as a query in search against HGSC data for *A. castellanii* Neff (http://www.hgsc.bcm.tmc.edu/projects/microbial/microbial-pubPreview.xsp?project_id=163 [last accessed 2011 Dec 18]). Identified contigs were searched for splicing sites to determine the gene intron/exon structures.

Construction of VDAC, Tom40, and Sam50/Tob55 Protein Sequences Datasets

Using NCBI Entrez (http://www.ncbi.nlm.nih.gov/sites/entrez [last accessed 2011 Dec 18]) and EBI SRS (http://srs.ebi.ac.uk/ [last accessed 2011 Dec 18]), we scanned available databases with a keyword search approach to find well annotated protein sequences of Sam50/Tob55, Tom40, and VDAC from species representing various eukaryotic lineages. In the next step, a set of known and well annotated Sam50/

Tob55, Tom40, and VDAC sequences (i.e., from *H. sapiens*, *M. musculus*, *A. thaliana*, *O. sativa*, *C. reinhardi*, *S. cerevisiae*, *Homo sapiens*, *Mus musculus*, *Arabidopsis thalliana*, *Oryza sativa*, *Chlamydomonas reinhardi*, *Saccharomyces cerevisiae*, *Neurospora crassa*, *Danio rerio*, *Drosophila melanogaster*, and *Caenorhabditis elegans*) were used as queries in Blast, TBlastN (Altschul et al. 1990), and PSIBlast (Altschul et al. 1997) searches to expand the data set with still unannotated sequences from other species. Two variants of parameters for Blast searches were used: 1) default ones with changes in the filtering and masking part—low complexity regions and mask for looking table only (soft masking) switched on and 2) word size—2, amino acid substitution matrix—BLOSUM45, gap opening cost—15, gap extension cost—2, and the same filtering options as above. The latter Blast variant was used to conduct more sensitive searches (finding the distant homologs). In the cases of multiple forms of a given protein (splice variants), a product of the longest transcript was selected. To identify VDAC, Tom40, and Sam50/Tob55 sequences in the genomes of microorganisms previously classified as protists which were still under the sequencing process, we searched against the TBestDB (http://tbestdb.bcm.umontreal.ca [last accessed 2011 Dec 18]) (*Allomyces macrogynus*, *Batrachochytrium dendrobatis*, *Capsaspora owczarzaki*, *Chlamydomonas incerta*, *Convoluta pulchra*, *Cyanophora paradoxa*, *Glaucocustis nostoineraum*, *Hyperamoeba* sp., *Isochryis galbana*, *Malawimonas californiana*, *Malawimonas jakobiformis*, *Mesostigma viride*, *Micromonas* sp., *Pavlova lutheri*, *Polysphondylium pallidum*, *Polytomella parva*, *Rhizopus oryzae*, *Saitoella complicata*, *Scenedesmus obliquus*, *Sphareoforma arctica*, *Spizelloyces punctatus*, and *Taphira deformans*) and the JGI Genome Projects resources (http://www.jgi.doe.gov/genome-projects/ [last accessed 2011 Dec 18]) (*Aureococcus anophagefferens*, *Chlorella* sp., *Dictyostellium purpurateum*, *Emiliania huxleyi*, *Naegleria gruberi*, *Ostreococcus* sp., *Phaeodactylum tricornutum*, *Phytophthora ramorum*, *Thalassiosira pseudonana*, and *Volvox carteri*) using TBlastN and BlastP algorithms, respectively.

During the final part of the data set collection process, we employed Hidden Markov models (Eddy 1998) that encoded protein domains identified in the analyzed β-barrel proteins: Porin3 domain (Pfam: PF01459) present in Tom40 and VDAC and Bac_surface_Ag (Pfam: PF01103) in Sam50/Tob55. Protein domains were identified with InterProScan (http://www.ebi.ac.uk/Tools/InterProScan/ [last accessed 2011 Dec 18], Quevillon et al. 2005), and the database search was performed with the HMMER package version 3.0b3 (Eddy 2009). We used NCBI nr database (nonredundant protein sequences) for the HMM search.

Multiple Sequence Alignment

Multiple sequence alignments of full-length VDAC, Tom40, and Sam50/Tob55 amino acid sequences were carried out with MAFFT version 6 (Katoh et al. 2002). To get more accurate alignments, we ran MAFFT with L-INS-i iterative refinement method, which is best suited for aligning sequences with one conserved domain and long gaps. Subsequently, the obtained alignments (Supplementary files 4, 6, and 8, Supplementary Material online) were subjected to T-COFFEE CORE (Notredame et al. 2000; Poirot et al. 2003) to identify and exclude poorly/incorrectly aligned regions. We used six as a CORE index score cut off value, which means that all regions with score below six were excluded from the phylogenetic analysis. The evaluated alignments (Supplementary files 5, 7, and 9, Supplementary Material online) were also manually checked and, when necessary, adjusted in JalView (Waterhouse et al. 2009).

Phylogenetic Analyses

Maximum likelihood (ML) analyses of the studied proteins were conducted using RAxML v7.2.8 (Stamatakis 2006) with RETRV amino acid substitution matrix. The matrix as well as the amino acids frequencies was estimated from input multiple sequence alignments. RAxML was executed with a rapid Bootstrapping algorithm (Stamatakis et al. 2008). The best tree topologies were determined under PROTCAT with a number of 25 per site rate categories. Support values for trees were assessed with 1,000 bootstrap replicates and the likelihood values of the final trees were evaluated and optimized under PROTGAMMA with four gamma-distributed discrete rate categories.

Phylogenetic trees obtained in the steps described above were visualized with Interactive Tree of Life (iTOL, http://itol.embl.de [last accessed 2011 Dec 18], Letunic and Bork 2007), midpoint rooted, and colored to visualize the respective phylogenetic clade membership of their terminal nodes.

Search for Cladistic Markers

Cladistic markers were selected arbitrary in a given position of full-length alignments of amino acid sequences of Tom40, VDAC, and Sam50/Tob55 (see supplementary material, Supplementary Material online). We required a given amino acid to be present at the same position in more than 80% of representative sequences from a given clade and below 20% in other clades. Predictions of the secondary structure of β-barrel proteins, essential to estimate the localization of the cladistic markers were conducted with the PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/ [last accessed 2011 Dec 18], Bryson et al. 2005).

Results

Identification of *A. castellanii*'s VDAC, Tom40, and Sam50/Tob55

Advances in molecular techniques and the accumulation of large amounts of sequences from genome sequencing projects and EST databases clearly allow for a comprehensive

VDAC

MSYAPAKPVSVPVVVVPTVVRPAAPLYSNLGKGAKDL**LSKGFPSTYKVEVTTSAENGVQFVS**
SAEKKQANKTDVVVGTFQPKYKLASRGLELTGTFDTDNQIKAELALDNLFVPGVKGTFKAQT
GASHDLEAAFEYKHEVGTFTSTFVHNPTTAKTLLSATATVSRQSVTAGVESKYSLAPAQPGT
LTTVTGALNYKAATHDLTAFVKSDAGSVTDGGDAPRLYSLGAHLHYTPSKESSFASSLDYDL
QKNTIKVTIGGSQKLDARTETKAKFSSDGRLGLGVAQQLTPAVKATLGTELNTFDLAGSTPK
FGVHFEVRA

Tom40

MDKLPSPPAAAPSALPSVPVMPPLPPLAGAGSP**SEPIAPAETHKEAPPSFWRRVLDHGPLPY**
PGKFDDLSKESKGILDSRENFDGFNFDVSRTITANFDMTHAIHFGSQTEPPSYNFMTRYFSP
TGMFLDGKLSPGPMLLNGTVQGRIAFPLKTWYRTTPGGRSKPVSKLSLRLTGVVGKPSAKPI
PPQFNMELDMKGSDYQAQFKWHNPGMYELSYAQSVHPNVALGLHGLFDSRMGRTELTAAGRV
AWNGRKCIATASMASYGHLTSSYTQVMEDGRKSWSTDMQLIHLEDHGVWEAVYSAGYAYNLL
QSRVRGRVDSNWCCSAILEEKVLDTVTLLICGDMNYAKDVYKVGFGFSFHL

Sam50/Tob55

MLGVRGEDAWRVRRVWVKGNRRTRPDVVAACVKPVLKARTFDEVLARVAEAAGELKGLGIFK
SVNLVLDDLPDDVAAADPSACDLRVELVEHKLTRIEVKTSTTVG**ENEPDVQATVGLCNAFGR**
AETVSVSAQVGASNWREHWSSAFSLAFKRPVLGQGPGPRHIEADATRASHRLPWCALTQTTH
GLALRYSLPSHQVSYEASARHVIPESSAPLALRSHAGHSLKSAVKYVYAHDTRDDPLTPTTG
HAFTSSTNELRELAGLGGDVRFVKQEMAAQLNLPLGKTRCSFNVLAKAGYLHALDSGSRIVD
RFFLGGPSIRGFQYNAVGPSHQQRALGGGAYWAGSLHLSFPLPLRDVPDFVSGHLFANAGNL
RQPTP

Fig. 1.—Identification of amino acid sequences of *Acanthamoeba castellanii* VDAC, Tom40, and Sam50/Tob55. The peptide sequences obtained by LC-MS/MS (marked with red letters) match amino acid sequences derived from *A. castellanii* EST database (bold letters). In the case of VDAC and Sam50/Tob55, also manually extended amino acid sequences (marked with underlined letters) were partially confirmed by LC-MS/MS (red and underlined letters). In each case, the rest of the sequence (regular font) was obtained by an analysis of contigs from the *A. castellanii* HGSC database and partially confirmed by LC-MS/MS analysis (red letters).

analysis. In case of an incomplete genome sequence, like the one of *A. castellanii*, there is an opportunity to apply mass spectrometry for protein analysis and identification. Therefore, we used *A. castellanii* EST data (O'Brien et al. 2007) as well as raw genome sequencing data to obtain amino acid sequences suitable for LC-MS/MS analysis (see Materials and Methods). We identified three EST clusters annotated as fragments of genes encoding the respective proteins. ACL00009286 (Sam50/Tob55) contained only one EST, whereas ACL00002106 (Tom40) and ACL00001709 (VDAC) contained four and three overlapping ESTs, respectively. These contigs were used as queries for Blast searches against the nonredundant protein database (nr) that resulted in multiple hits to well-known sequences of the studied proteins (28–32%). Consequently, the translated EST contigs were used as the source for the analysis of mass spectrometry results.

Mitochondrial outer membrane proteins of *A. castellanii* with MWs close to those determined for known VDAC (about 36 kD), Tom40 (about 41 kD), and Sam50/Tob55 (about 48 kD) were subjected to LC-MS/MS. Peptides obtained by LC-MS/MS closely matched the reference sequences with 127 identical residues of 140 amino acids in total for VDAC, 117/279 for Tom40, and 79/128 for Sam50/Tob55 (see fig. 1). Thus, mass spectrometry results strongly confirmed the EST data, and it can be concluded that the deduced partial sequences of the studied *A. castellanii* proteins are correct. Afterward we extended nucleotide sequences by collecting traces, assembly, and basic Blast analysis. We were able to extend the VDAC and Sam50/Tob55 sequences (underlined letters in fig. 1) but not the Tom40 sequence. The amino acid sequences were extended from 140 to 319 residues for VDAC and from 128 to 376 for Sam50/Tob55 and were confirmed by ESTs. The added peptides were partially confirmed by LC-MC/MS analysis, that is, 52 amino acids for VDAC and 32 amino acids for Sam50/Tob55 (fig. 1, red and underlined letters). In each case, the rest of the sequence was obtained by an analysis of contigs from the *A. castellanii* HGSC database and partially confirmed by LC-MS/MS analysis (red letters in fig. 1).

Results of the search against the *A. castellanii* genome indicated that in the case of TOM40, we were able to identify a complete ORF coding for 361 amino acid long protein that was confirmed by standard PCR (5′ end sequence) and 3′ RACE (3′ end sequence) analysis. 3′ RACE resulted in two products of predicted 1100 and 1000 bp size (see supplementary fig. 1, Supplementary Material online). Moreover standard PCR of 5′ end sequence of TOM40 resulted in product of predicted size of 934 bp. The sequences of 934 and 1000 bp products were verified by sequencing. *Acanthamoeba castellanii* mRNA sequence of TOM40 was deposited in the GenBank under accession number HQ259301. For Sam50/Tob55, we obtained 376 amino acids and 319 for VDAC (bioinformatic analysis partially confirmed by LC-MS/MS, red letters, fig. 1). Secondary structure analysis strongly suggested that each of the protein

encoded by identified sequences is able to form transmembrane β-strands, which is a crucial feature of membrane β-barrel proteins. Finally, InterProScan search revealed a Porin3 domain (Pfam: PF01459) to be present in the inferred *A. castellanii* Tom40 (amino acids 63–354) and VDAC (amino acids 25–312) protein sequences and Bac_surface_Ag (Pfam: PF01103) in Sam50/Tob55 (amino acids 120–371). Thus, the obtained results constitute a strong line of evidence that the deduced sequences are indeed part of *A. castellanii* VDAC and TOM and SAM/TOB complexes.

Identification of *D. discoideum* Sam50/Tob55

Although, the *D. discoideum* genome project was completed in 2005 (Eichinger et al. 2005), only two genes encoding mitochondrial outer membrane β-barrel channels were characterized in this genome, namely genes for Tom40 (Maćasev et al. 2004) and VDAC (Troll et al. 1992). For identification of the remaining Sam50/Tob55, we applied a bioinformatic approach. Sam50/Tob55 protein sequences from fungi, animals, and plants were used as Blast queries against the *D. discoideum* AX4 genome (Eichinger et al. 2005) to search for homologous sequences. The best matching sequence was found on chromosome 1. It was interrupted by two introns and potentially encoded for a 396 amino acid long protein with a MW of 44.6 kDa. This sequence has already been deposited at the dictyBase database (http://dictybase.org [last accessed 2011 Dec 18]) under DDB_G0269550 and DDB0305221 Gene ID and Sequence ID, respectively. The amino acid sequence presented greater similarity to animal (e.g., bee, human, chicken, or clawed frog) than to fungal or plant Sam50/Tob55 sequences, sharing 21–23% sequence identity. The predicted protein sequence contained the Bac_surface_Ag protein domain (Bacterial Surface Antigen, Pfam: PF01103) spanning between position 96 and 392. Because we were able to find this domain in all well annotated animal, fungal, or plant Sam50/Tob55 sequences, the finding provided additional support that the described sequence is a true ortholog of Sam50/Tob55. Moreover, a reciprocal BlastP search against the nr database did not produce any significant nontarget hits.

Analysis of Evolutionary History of Mitochondrial Outer Membrane Channels

As mentioned above, VDAC, Tom40, and Sam50/Tob55 are proteins of the mitochondrial outer membrane that display channel activity. They are also termed mitochondrial β-barrel proteins as they have a β-barrel topology similar to proteins present in the outer membrane of Gram-negative bacteria (Walther et al. 2009). The other β-barrel proteins, Mdm10 and Mmm2, also located in the outer membrane of mitochondria (Walther et al. 2009; Yamano et al. 2010), have been excluded from our analysis, as they are not found outside the fungal lineage. Sam50/Tob55 is homologous to the bacterial Omp85 outer membrane protein, which confirms the endosymbiotic origin of mitochondria. The other two proteins, Tom40 and VDAC, do not share significant sequence similarity with bacterial outer membrane proteins and probably evolved later in the course of evolution of mitochondria from endosymbiont. However, they have been found along the whole eukaryotic lineage. Recently, it has been shown that mitochondrial β-barrels belong to two different groups: the 19-stranded barrels (Tom40 and VDAC) and 16-stranded barrels (Sam50/Tob55) (Zeth 2010), which may indeed suggest their independent origin.

We decided to use Sam50/Tob55, Tom40, and VDAC in the phylogenetic analysis because they present a high conservation in their function as mitochondrial outer membrane channels. Compared with other similar studies (e.g., Young et al. 2007), we focused not only on animals, fungi, or plants but also extended our initial data set with Sam50/Tob55, Tom40, and VDAC sequences that we identified in many unicellular eukaryotes, including the sequences identified for *A. castellanii* and *D. discoideum*. As shown in table 1 and in details in supplementary table 1, Supplementary Material online, we analyzed the mitochondrial outer membrane β-barrels from 157 species, including 7 Amoebozoas, 89 Opisthokontas, 32 Archaeplastidas, 20 Chromalveolatas, 8 Excavatas, and a single Rhizaria species. The obtained data sets consisted of 84, 124, and 184 protein sequences of Sam50/Tob55, Tom40, and VDAC, respectively (see supplementary files 1–3, Supplementary Material online), and were used to construct phylogenetic trees.

Phylogeny of VDAC

As a result of gene duplication events during the evolution of the VDAC protein family, VDAC, the most ubiquitous protein located in the outer membrane of mitochondria (Gonçalves et al. 2007; Shoshan-Barmatz et al. 2010), has multiple genes (isoforms) found in numerous species. For example, in vertebrates the VDAC family consists of three paralogous genes (Saccone et al. 2003; Young et al. 2007), but in plant genomes, up to five paralogs of VDAC were identified (Wandrey et al. 2004). The latter can be explained by the very common phenomenon of polyploidy in plants rather than being ascribed to evolutionary pressure to create more specialized forms of VDAC, as it is observed in mammals (De Pinto et al. 2010). Usually, VDAC paralogous genes are located on different chromosomes, with the exception of *D. melanogaster* genome where genes for two VDAC isoforms are located on the same chromosome (2L) in a very close proximity and are not separated by any other gene suggesting tandem duplication of the *VDAC* gene.

A phylogenetic tree of 184 VDAC amino acid sequences from 123 species representing five eukaryotic supergroups (Opisthokonta, Archaeplastida, Amoeboza, Chromalveolata,

Table 1
Summary of Distribution of Mitochondrial Outer Membrane β-Barrels Used in This Study among the Six Eukaryotic Supergroups

Supergroups	Rank	No. of Species	Sam50/Tob55	Tom40	VDAC
Amoebozoa	Eumycetozoa	4	2	3	4
	Acanthamoebidae	1	1	1	1
	Entamoebida	2	2	2	—
Opisthokonta	Animals	44	28	43	76
	Choanomonada (Choanoflagellata)	2	1	1	1
	Fungi	41	30	34	31
	Mesomycetozoa	2	—	1	2
Archaeplastida	Chloroplastida	30	12	16	48
	Glaucophyta	2	—	—	3
Chromalveolata	Haptophyta	3	1	2	3
	Stramenopiles	4	5	4	4
	Alveolata	12	1	12	4
	Cryptophyceae	1	—	1	—
Excavata	Malawimonas	2	—	—	2
	Parabasalia	1	—	1	1
	Preaxostyla	1	—	1	—
	Heterolobosea	1	1	1	1
	Euglenozoa	3	—	—	3
Rhizaria	Cercozoa	1	—	1	—
Total		157	84	124	184

and Excatava; no sequences of the Rhizaria group were identified) was constructed using ML method (fig. 2). Two large eukaryotic supergroups—Opisthokonta (including Metazoa, Fungi, Choanomonada, and Mesomycetozoa) and Archaeplastida (including plants) form monophyletic clades. However, only animal and fungal groupings are supported by high, that is, over 80%, bootstrap values. Discussed above VDAC paralogs are represented by subclusters within animal and plant clades. *Acanthamoeba castellanii*, *D. discoideum*, and other representatives of Amoebozoa are grouped together, though they are weakly supported by bootstrap values. Interestingly, excavate and alveolate sequences do not cluster into monophyletic groups. Unfortunately, many of the internal nodes, that are most essential for this analysis, are supported by low bootstrap values. Consequently, the phylogenetic relationships could not be inferred with high confidence, which is most likely caused by a high sequence divergence.

Phylogeny of Sam50/Tob55

Sam50/Tob55 is an essential member of the SAM/TOB complex responsible for the assembly and insertion of β-barrel precursor proteins into the mitochondrial outer membrane as well as biogenesis of the TOM complex (Dolezal et al. 2006; Bohnert et al. 2007; Neupert and Herrmann 2007; Becker et al. 2008; Walther et al. 2009). It is known to be the only member of mitochondrial protein import machinery of the outer membrane that is homologous to bacterial β-barrel proteins (Omp85) and is also conserved along the whole eukaryotic lineage (Walther et al. 2009).

A data set of 84 amino acid sequences from 77 species was used to infer the phylogenetic relationships of Sam50/Tob55 (5 representatives of Amoebozoa, 56 of Opisthokonta, 9 of Archaeplastida, 6 of Chromalveolata, and 1 of Excavata). Generally, with a few exceptions in plants, for example, *A. thaliana* and *O. sativa japonica*, and fishes, for example, *D. rerio* and *Tetraodon viridis*, where observed paralogs were lineage-specific studied genomes contained a single copy of the gene encoding Sam50/Tob55. Figure 3 presents a phylogenetic tree of Sam50/Tob55 amino acid sequences constructed with the ML approach. Opisthokonta and Archaeplastida supergroups form monophyletic clades. Although, animal, fungal, and plant sequences group into three distinguishable clades, each of them united by a single node, these clades are not statistically supported. The relationships of the remaining eukaryotic supergroups in the Sam50/Tob55 data set, that is, Amoebozoa, Chromalveolata, and Excavata, could not be resolved with significant support values. Similar to the VDAC data set, whereas the external nodes of the Sam50/Tob55 tree are well supported by bootstrap values, the internal nodes that connect different phylogenetic groups present low bootstrap values with no statistical significance.

Phylogeny of Tom40

Tom40 is the most important unit of the TOM complex (e.g., Dolezal et al. 2006; Bohnert et al. 2007; Neupert and Herrmann 2007; Becker et al. 2008; Walther et al. 2009). This integral outer mitochondrial membrane protein creates channels that are crucial for protein import into mitochondria. Tom40, together with components like Tom7 and Tom22, forms the main core of the TOM complex, which has homologous sequences found in a wide range of eukaryotic organisms (Perry et al. 2008).

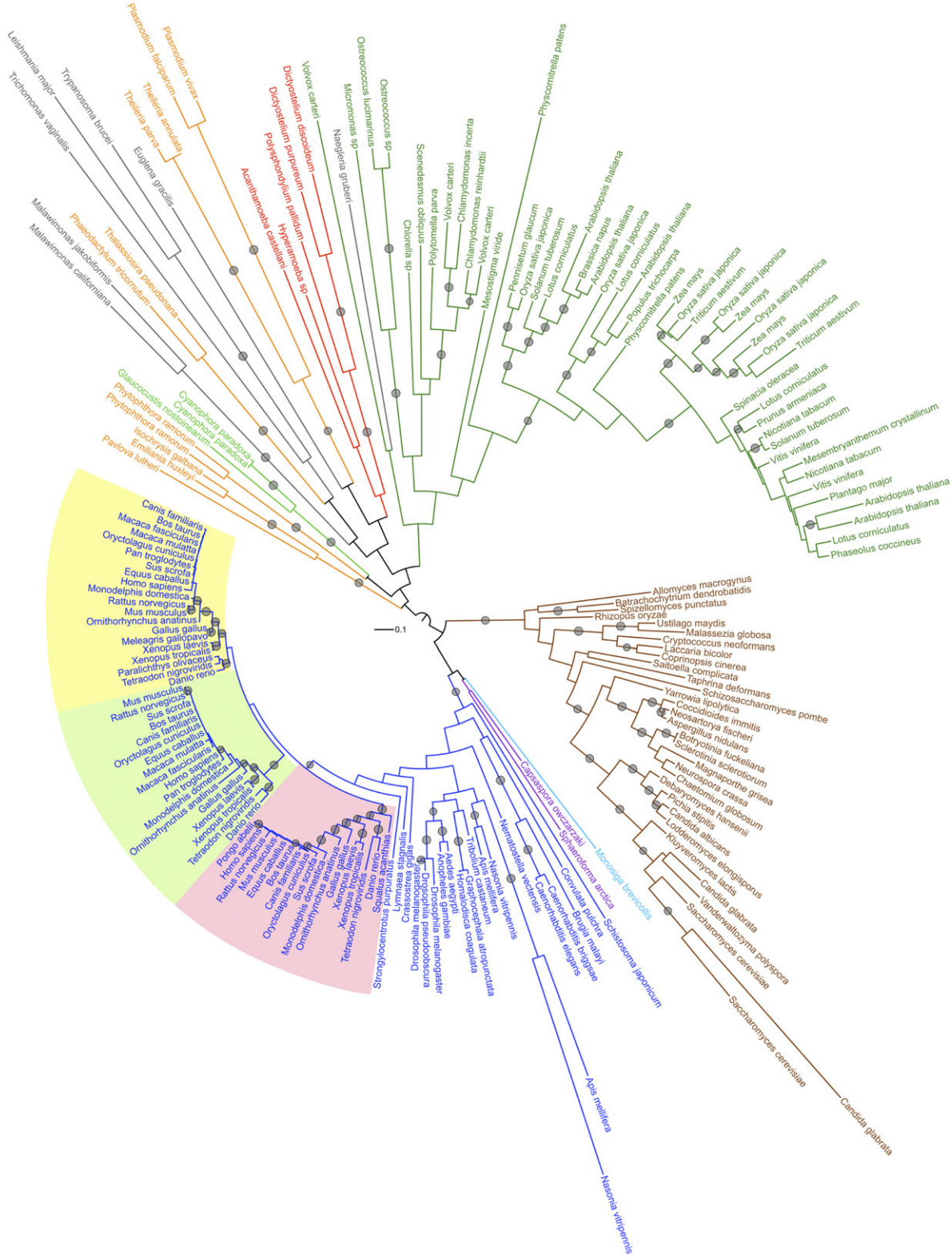

Fig. 2.—Evolutionary relationships of 184 VDAC protein sequences based on ML analysis. Color key: red—Amoebozoa, blue—Metazoa (Opisthokonta), light blue—Choanomonada (Opisthokonta), violet—Mesomycetozoa (Opisthokonta), brown—Fungi (Opisthokonta), green—Chloroplastida (Archaeplastida), light green—Glaucophyta (Archaeplastida), orange—Chromalveolata, gray—Excavata, and yellow—Rhizaria. Colored leaf ranges denote three subclades created by three vertebrate VDAC isoforms. Filled circles represent bootstrap values of 80% and above.

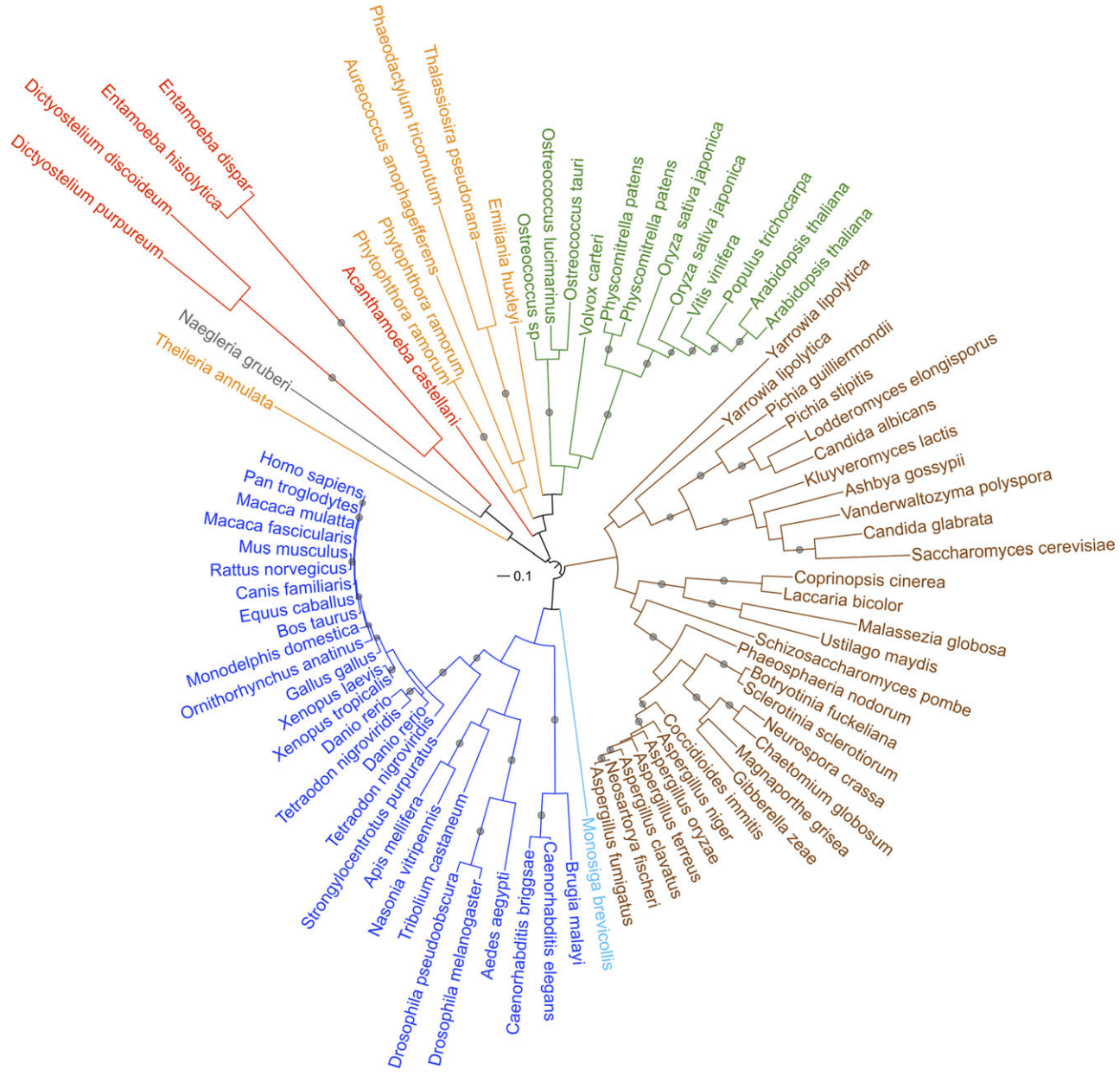

Fig. 3.—ML tree of 84 Sam50/Tob55 protein sequences. Colors as described in figure 2. Filled circles represent bootstrap values of 80% and above.

The data set used to infer phylogenetic relationships of Tom40 consisted of 124 amino acid sequences identified in 107 species representing all six eukaryotic supergroups. The ML method was used to resolve phylogenetic relationships (see fig. 4). As in the case of VDAC and Sam50/Tob55, a similar arrangement of clades emerges for Tom40 sequences. Two of the six eukaryotic supergroups, namely Opisthokonta and Archaeplastida, form monophyletic groupings. However, within the first group, only the animal clade received a reliable statistical confidence, and Opisthokonta as a whole was not resolved with enough statistical power. Furthermore, our preliminary results indicated that *A. castellanii*'s TOM complex is rather of the animal type (Wojtkowska et al. 2005)—a protein complex identified electrophysiologically as the *A. castellanii* TOM complex contained a homolog of yeast/animal Tom70. Its MW (about 500 kDa) also pointed *A. castellanii* to an evolutionary relation with fungi and animals. In the present study, Tom40 sequence of the discussed amoeba is located closely to the plant node; however, its position is not definite considering the obtained bootstrap value. Similarly to VDAC and Sam50/Tob55 trees, the internal nodes at the base of the supergroups are not supported by bootstrap analysis, whereas the external ones that group closely related species

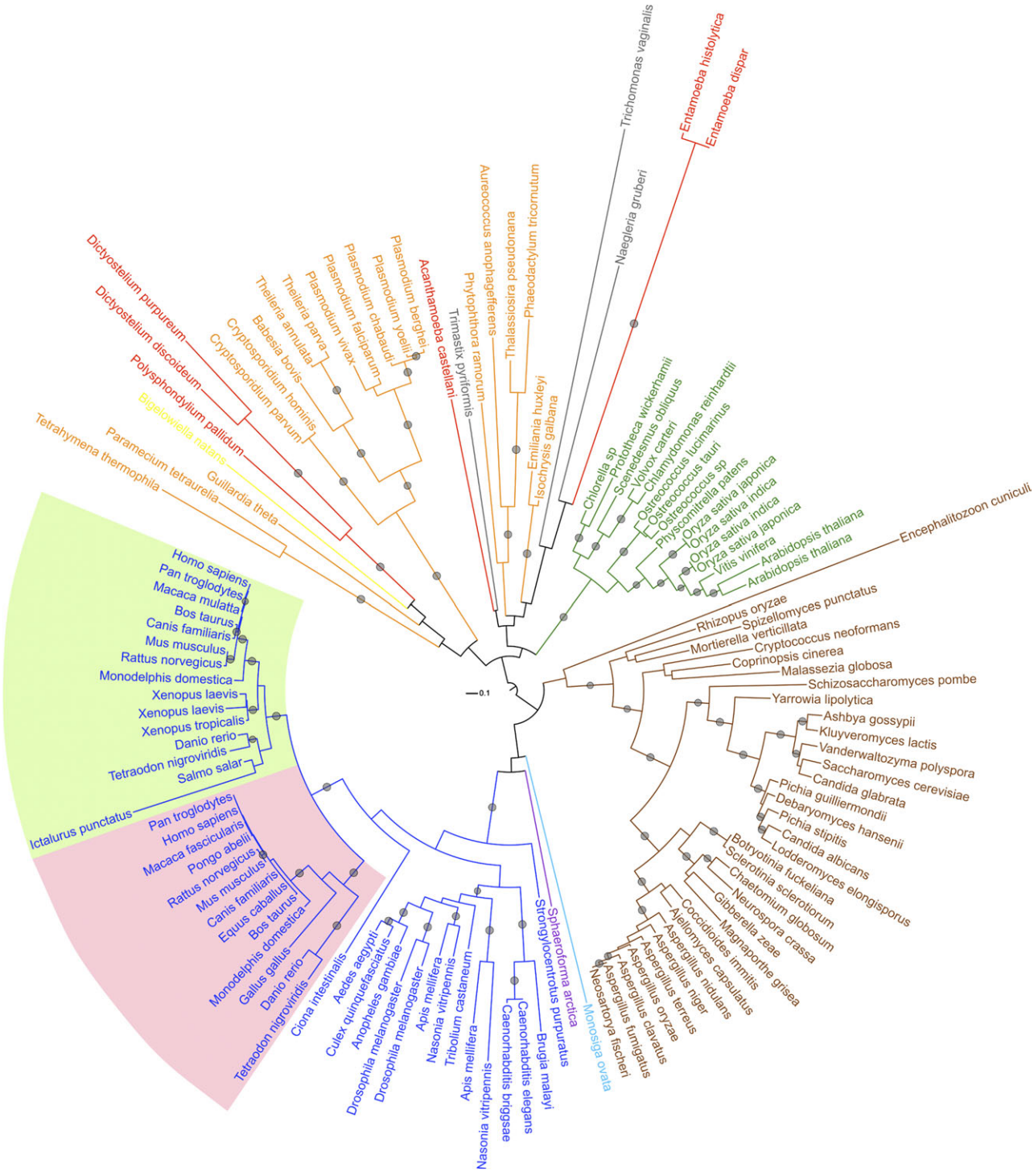

Fig. 4.—ML tree of 124 Tom40 amino acid sequences. Colors as described in figure 2. Colored leaf ranges denote two subclades created by two vertebrate Tom40 isoforms. Filled circles represent bootstrap values of 80% and above.

are well supported. Therefore, it is impossible to draw any definitive conclusions on major groups relationships.

One can clearly notice that in the presented phylogenetic tree, there are two vertebrate paralogs of Tom40. The second Tom40 form has been previously reported by Kinoshita et al. (2007) in the rat genome along with its human and mouse orthologs. The mammalian Tom40 isoforms are known as Tom40A and Tom40B or Tom40 and Tom40L

and are encoded by *TOMM40* and *TOMM40B* genes, respectively, and located on different chromosomes (Kinoshita et al. 2007). The second form is slightly shorter, causing an N-terminal truncation. It shares about 60–65% of protein identity with the "primary" form (pairwise alignments of mammalian sequences). The comparison of the exon/intron structure (supplementary fig. 2, Supplementary Material online) reveals nine exons to be present in both forms (exception: ten exons in *TOMM40B* of opossum). Our studies of the chromosomal location and the gene vicinity of both forms performed on many species did not indicate any segmental duplication. Most likely, only this single gene was duplicated in the early vertebrate evolution, resulting in the observed paralogs. Two possible hypotheses can explain the existence of two Tom40 sequences in vertebrate genomes. In the first scenario, both isoforms are located within the same complex and function as pores (Kinoshita et al. 2007). The following assumption is consistent with an observation that at least two channels are located in a single TOM complex what has been verified by electron micrograph studies (Ahting et al. 1999). In the second scenario, there are two separate types of the TOM complexes present in the mitochondrial outer membrane with channels formed either by Tom40 or by Tom40L (Kinoshita et al. 2007). In this case, both complexes cooperate with each other in protein import into mitochondria.

Interestingly, the existence of two Tom40 isoforms in vertebrates encoded by separate genes resembles the presence of three vertebrate VDAC isoforms, also encoded by distinct genes. It has been suggested that distinct VDAC genes found in vertebrates follow a clock-like behavior, thus suggesting that new and specific, probably indispensable functions, have been generated by gene duplication events (Saccone et al. 2003). Accordingly, VDAC and Tom40 belong to the same group of β-barrel proteins and they both participate in metabolite transport (Antos et al. 2001; Budzińska et al. 2009). These common characteristics of function and structure might be caused by similar evolutionary events.

Cladistic Markers for VDAC, Sam50/Tob55, and Tom40

The bootstrap values for some clades in the trees presented in figures 2–4 are relatively low. To improve trees reliability, we searched for unique and common amino acid residues in sequence alignments that could serve as cladistic markers (table 2). The analysis of amino acid sequence alignments of VDAC, Sam50/Tob55, and Tom40 resulted in cladistic markers only for subclades of animals and fungi or supergroups like Opisthokonta and plants. Interestingly, the highest number of markers was found for VDAC, lesser for Tom40, and the smallest amount for Sam50/Tob55. In the case of Sam50/Tob55, only one putative marker was identified, that is, aspartic acid at position 190 in animal sequences distinguishing them from other systematic groups. For Tom40, asparagine at position 471 and proline at position 472 distinguish Opisthokonta from Archaeplastida, Chromalveolata, Amoebozoa, and Excavata. On the other hand, plant Tom40 could be discriminated from representatives of other clades by two markers—glutamine at position 222 and valine at position 147. Furthermore, fungal Tom40 could be distinguished by proline at position 222, whereas animal Tom40 contains threonine at position 468 of the alignment. For VDAC, we have observed ten markers in the case of the animal sequences and some of them are present within four characteristic fragments that differ in length (underlined sequences in table 2). There are also six possible markers for Opisthokonta and four of them are located within one common fragment containing 14 amino acids. We also identified two possible markers for plants and one for fungi.

Although we did not find any cladistic markers for Amoebozoa as a separate clade, we searched for common residues between aligned amino acid sequences of the analyzed clades and *A. castellanii* (table 2). For Tom40, we did not find such common residues, but for Sam50/Tob55, we found one residue that is common with animals (aspartic acid at position 190), and in the case of VDAC, there are two common residues with Opisthokonta and four with animals. Next, we analyzed localization of the putative markers in the predicted protein secondary structures (table 2). In the case of VDAC, they were mostly localized in β-strands (b3, b7/8, b12, b13, b14, b17, b18, and b19) and within N-terminus between α-helix and the first β-strand. In the case of Tom40, they are localized in β-strands (b13 and b8), between β-strands (b9 and b10), or within the N-terminus. Interestingly, in the case of Sam50/Tob55, the possible cladistic markers were localized within the POTRA domain (positions 62 and 140). Thus, only in the case of Sam50/Tob55, the possible cladistic markers were mostly localized within known conserved domains (Zeth 2010; Zeth and Thein 2010). According to authors, these are β-strands 2–4 and 18–19 for Tom40, β-strands 18–19 for VDAC, and β-strands 12–16 and interface between the β-barrel and the POTRA domain for Sam50/Tob55.

Discussion

In the last decade, a new system that takes into account the strong differentiation of representatives of the former kingdom of Protista has slowly replaced the traditional classification of Eukaryotes consisting of four kingdoms. Unlike the previous one, the new system is based not only on morphological but also on molecular data and divides eukaryotic organisms into six supergroups (Adl et al. 2005; Keeling et al. 2005). According to this system, the subjects of this study, *A. castellanii* and *D. discoideum*, are now classified under the supergroup of Amoebozoa located basally to the Opisthokonta supergroup, which includes animals and fungi. On the basis of the roles that the mitochondrial

Table 2
Summary of the Molecular Markers Identified for Tom40, Sam50/Tob55, and VDAC

Protein	Clade	Position	AC	Marker	In Sequence	Domain
Tom40	Animals	468		T	AVTLGNPD	b8
	Fungi	222		P	PLFQVSH	b2
	Opisthokonta	471		N		
		472		P	AVTLGNPD	b9
	Plants	147		V	DYSNLPCPVLYE	N-terminus
		222		Q	NKSLNQKFFLSH	b13
Sam50	Animals	190	+	D	RTKDDII	POTRA
VDAC	Animals	95	+	G		Between α-helix and b1
		103		F	DVFNKGYGFGLV	
		139	+	T	SSNTDTGVKSGS	b3
		226		K		
		241		D	KIKTSYKRECINLGCD	Between b7 and b8
		349		H		
		354		V	DFQLHTNCVN	b12
		372	+	G	DGTEFGGSIYQKV	b13
		392		T	KLETAVNLAWTA	b14
		493	+	G	LIGLGYTQTLRPG	
		525		L		
		540		G	VKLTLSALVDGKNFNAG	b18
		586		E	KLGLGLELEPSA	b19
	Fungi	239		R	NFHARAFFDL	b8
	Opisthokonta	165	+	T		
		166		Q		
		168		W		
		172	+	N	TQKWNTDNTLGTEI	Between b4 and b5
		201		P	SSFSPNTGKKN	Between b6 and b7
		570		H	GGHKLGLGLELE	b19
	Plants	165		D	TVIDVKVDTD	b3
		495		L	GKASALIQHEWRPK	b17

NOTE.—Cladistic markers were selected arbitrary in a given position if a given amino acid was present with a frequency higher than 80% in a given clade, whereas in others, the frequency was below 20%. "b" denotes a given β-strand structure in reference to the secondary structure of the human Tom40, Sam50/Tob55, and VDAC 1 (as described in Zeth 2010). "+" in the AC column denotes presence of a marker in a sequence of *Acanthamoeba castellanii* protein.

outer membrane channels of a predicted β-barrel topology have in the functioning and biogenesis of mitochondria, we hypothesized that their analysis in these and other representatives of Amoebozoa should contribute important data concerning proposed systematic positions of this phylogenetic group.

However, it should be taken into account that the interactions between β-barrel channels and other subunits within the TOM and the SAM/TOB complexes allow the communication and exchange between mitochondria and the surrounding cytosol which forms the platform for a variety of protein–protein interactions (Zeth 2010). Therefore, a secondary loss of the appended components of the TOM and SAM/TOB complexes could sometimes result in differences in mitochondrial β-barrel proteins. For example, parasitic organisms, like *Entamoeba histolytica* (Tovar et al. 1999; Dolezal et al. 2010) or *Encephalizoon cuniculi* (Waller et al. 2009) that contain mitochondrion-related relict organelles termed mitosomes, seem to be missing almost all the additional components of the TOM and SAM/TOB complexes what is caused by the reductive evolution affecting mitosomal organelle. The same may even apply to VDAC. Though, it forms a channel as a monomer and it was also reported to form big homo- and heterocomplexes (e.g., Shoshan-Barmatz et al. 2010) with additional functional subunits that could have been either gained or lost.

In contrast to animals or plants, where duplication events took place during the course of evolution, the analyzed Amoebozoa species, including *A. castellanii* and *D. discoideum*, have only single genes coding for VDAC, Tom40, and Sam50/Tob55. However, even within this supergroup, the available sequences appear to evolve quickly. For example, when the corresponding amino acid sequences from *A. castellanii* and *D. discoideum* were compared, they yielded sequence similarity values not higher than 55% (pairwise alignments with different alignment tools). This is reflected in the presented trees where *A. castellanii* never groups next to *D. discoideum* but is rather in a basal position to all amoebozoans (e.g., VDAC tree in fig. 2), what might suggest a higher rate of amino acid substitution in mitochondrial β-barrels in this lineage. Moreover, *A. castellanii*, although it lives free in soil or fresh water, can also function as a parasite causing, for example, encephalitis in human (Visvesvara 2010). This might suggest that its sequences

could have undergone specific changes including degeneration as a consequence of an adaptation to a parasitic lifestyle. Also other nonamoebozoan microorganisms analyzed here might have undergone similar changes and their mitochondrial outer membrane β-barrel channels could have been co-opted for new functions.

The presented phylogenetic trees also demonstrate that the number of analyzed protein sequences does not allow to perform a fully comprehensive survey of the evolutionary history of the mitochondrial outer membrane β-barrels of Amoebozoa species. This also pertains to the representatives of Excavata and Rhizaria, and partially to Chromalveolata, and it seems to be caused by the number of genomic sequences that are currently available for unicellular organisms previously referred to as protists. In case of abundantly sequenced representatives of Opisthokonta and Archaeplastida, the obtained results not only match the proposed system of eukaryotic classification but also show that in these lineages paralogous genes have emerged a couple of times in the evolution of mitochondrial β-barrel channels. However, for Amoebozoa, Chromalveolata, Excavata, and Rhizaria, it might be difficult to "attract" sequences of other organisms of the same phylogenetic group to form strong and well separated clades that could be satisfyingly supported in phylogenetic trees. On the other hand, the protein sequences of the outer membrane β-barrels of the former Protista kingdom might have diverged in such a way that, by chance, they are presently more similar to those from other eukaryotic supergroups. Nevertheless, the obtained trees concerning Amoebozoa, Chromalveolata, Excavata, and Rhizaria point to some regularity. Sequences of *A. castellanii*, *D. discoideum*, and other amoebozoan mitochondrial outer membrane β-barrels cluster together with representatives of Chromalveolata and Excavata, in a close proximity to Archaeplastida. Thereby, in the outcomes of our phylogenetic studies, the former protist species do not follow the pattern of distribution proposed in the six supergroups classification system. Contrary, cladistic markers point at clustering of *A. castellanii* with animals and other opisthokonts what supports the proposed classification system. Thus, cladistic markers could be applied as an alternative approach to classify proteins in the phylogenetic clades that are not supported by the high bootstrap values (Das et al. 2008).

The improvement of metabolite transport as well as formation of mitochondrial protein import apparatus undoubtedly played a crucial role in the evolution of eukaryotic cell. The advances are especially noticeable within the outer membrane that serves as the "mitochondrial gate" for proteins and metabolites. This is probably caused by and is coupled with the evolution of lineage-specific additional components of the mitochondrial outer membrane import machinery and metabolite transport systems, which partially took over some functions of the very first simple transporters. It is likely that for functional and structural variability of proteins targeted to mitochondria, the development of additional subunits of the import machinery was evolutionarily beneficial. This is observed, for instance, in the case of the SAM/TOB complex: plants and animals contain metaxins that function as the receptor subunits, whereas in the fungal lineages Sam35/Tob38 and Sam37/Mas37 serve the same purpose (Chan and Lithgow 2008). Integral β-barrels of the mitochondrial outer membrane that are assembled with help of the SAM/TOB complex can have hydrophobic exterior domains or loops connecting individual β-strands. Their exposure to the cytosol would make the protein folding and/or insertion either difficult or impossible and that was likely one of the main reasons for the receptor subunits to evolve (Chan and Lithgow 2008). It is further speculated that, rather than Eukaryote-specific linear signals, a folding intermediate of a precursor with high β-sheet content serves as a signal for β-barrel proteins to assure their targeting from the cytosol to the outer membrane of mitochondria (Walther et al. 2009). It should also be remembered that the SAM/TOB complex is required for the assembly of precursors of the TOM complex, including not only the β-barrel subunit Tom40 but also a subset of α-helical subunits (e.g., Thornton et al. 2010).

As mentioned above, VDAC and Tom40 appear to share some properties, that is, predicted type of β-barrel structure or the presence of isoforms in vertebrates, which might explain Tom40's involvement in metabolite transport (Antos et al. 2001; Budzińska et al. 2009). The interactions of Tom40 with different subunits within the TOM complex are most likely conserved among species and are mainly based on the conserved patches localized on β-strands 2–4 and β-strands 18–19 (Zeth 2010). For VDAC, which is known to have different functions besides metabolite transport and is involved in various interactions with different proteins, the sequences seems to be more diverged and the identified significantly conserved region comprises β-strands 18–19 (Zeth 2010). The function of Sam50/Tob55 seems to be more specialized, as it is known to be involved only in protein import (e.g., Dolezal et al. 2006; Bohnert et al. 2007; Neupert and Herrmann 2007; Becker et al. 2008; Walther et al. 2009). The cross talk between Sam50/Tob55 and other subunits of the SAM/TOB complex is also likely conserved. The conserved segments of Sam50/Tob55 are localized on the β-strands 12–16 and interface between the β-barrel and POTRA domain (Zeth 2010).

Results of the present study indicate clearly that despite strong structure–function conservation, VDAC, Tom40, and Sam50/Tob55, diverged widely in the amino acid sequences. However, the number of cladistic markers found corroborates with the functional findings. Cladistic markers used to exam functional implication of the observed high sequence divergence appear to constitute amino acids essential for secondary structure and function of β-barrel channel. As shown in table 2, Sam50/Tob55 contains the lowest

number of possible cladistic markers, which most likely reflects its highly specialized function The higher number of cladistic markers observed for Tom40 resulted from more divergent function involving not only import but also metabolite transport as well as from the presence of isoforms. The highest number of possible cladistic markers for VDAC is caused presumably by the presence of different isoforms that may play different functions (Saccone et al. 2003; Young et al. 2007). However, it should be emphasized that no cladistic markers were obtained for Chromalveolata, Excavata, Rhizaria, and Amoebozoa, possibly due to small number of available sequences and due to their high sequence divergence.

In the case of VDAC, the evolving structure–function paradigm necessitates that the encoding gene redundancy should indicate a need for transport function innovation as well as the tendency to duplicate the genetic material as suggested for invertebrates, plants, and vertebrates (Saccone et al. 2003). Essentially, the same logic can be extrapolated to Tom40, and its involvement in proteins transport in vertebrates. In the case of vertebrate (chordate) VDAC isoforms, it is estimated that the divergence between VDAC3 (the oldest isoform) and VDAC1/2 (precursor for VDAC2 and VDAC1) occurred about 365 ± 60 Ma (Shoshan-Barmatz et al. 2010). Interestingly, at this time, formation of vertebrates probably took place, and it is believed that genome duplication was involved in the process (Saccone et al. 2003; Shoshan-Barmatz et al. 2010). Accordingly, our results indicate two isoforms of Tom40 in vertebrates and two isoforms of Sam50/Tob55 in fish lineage. It has been also reported that both VDAC and Tom40 contain 5 repeats of 50 amino acids motive that suggests common ancestry of two proteins (Zeth and Thein 2010).

In conclusion, many of the recently published works have already confirmed the monophyly of Amoebozoa (Burki et al. 2009; Hampl et al. 2009; Minge et al. 2009) or Excavata (Hampl et al. 2009), and in the case of Chromalveolata, the group was shown to be polyphyletic. A new clade was proposed, namely SAR, which is composed of two subclades of Chromalveolata, namely Stramenopiles and Alveolata, and the Rhizaria supergroup (Burki et al. 2007, 2008; Hackett et al. 2007). Polyphyletic character of Chromalveolata is also reflected in the pattern of evolution of mitochondrial β-barrel channels sampled from representatives of this supergroup and analyzed in this study. However, the statistical support obtained for many internal branches bearing on the validity of the supergroups is relatively poor, so the results of the current survey allow neither sustaining nor refuting the proposed system of classification. Nevertheless, given the constraints of sequencing and annotation for species of some of the supergroups analyzed here, the opportunities are open for further studies to delve deeper into their phylogenies and the history of mitochondrial outer membrane β-barrels within the various lineages.

Supplementary Material

Supplementary figures 1 and 2, files 1–9, and table 1 are available at the *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We are very grateful for the help received from Marcel Deponte and Walter Neupert. This project was supported by the Polish Ministry of Science and Higher Education (Grant N N303 143937) and the 6.FP Marie Curie Host Fellowships for the Transfer of Knowledge grant (FUNGEN). M.J. research was supported by the IoB funds. T.P. is supported by Intramural Research Program of the National Institutes of Health, U.S. National Library of Medicine.

Literature Cited

Adl SM, et al. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol. 52:399–451.

Ahting U, et al. 1999. The TOM core complex: the general protein import pore of the outer membrane of mitochondria. J Cell Biol. 147:959–968.

Alsam S, et al. 2006. Escherichia coli interactions with Acanthamoeba: a symbiosis with environmental and clinical implications. J Med Microbiol. 55:689–694.

Altschul SF, et al. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Anderson IJ, et al. 2005. Gene discovery in the Acanthamoeba castellanii genome. Protist 156:203–214.

Antos N, Budzińska M, Kmita H. 2001. An interplay between the TOM complex and porin isoforms in the yeast Saccharomyces cerevisiae mitochondria. FEBS Lett. 500:12–16.

Bay DC, Court DA. 2002. Origami in the outer membrane: the transmembrane arrangement of mitochondrial porins. Biochem Cell Biol. 80:551–562.

Bayrhuber M, et al. 2008. Structure of the human voltage-dependent anion channel. Proc Natl Acad Sci U S A. 105:15370–15375.

Becker T, Vögtle FN, Stojanovski D, Meisinger C. 2008. Sorting and assembly of mitochondrial outer membrane proteins. Biochim Biophys Acta. 1777:557–563.

Blachly-Dyson E, Forte M. 2001. VDAC channels. IUBMB Life 52:113–118.

Bohnert M, Pfanner N, van der Laan M. 2007. A dynamic machinery for import of mitochondrial precursor proteins. FEBS Lett. 581: 2802–2810.

Bryson K, et al. 2005. Protein structure prediction servers at University College London. Nucleic Acids Res. 33:W36–W38.

Budzińska M, et al. 2009. The TOM complex is involved in the release of superoxide anion from mitochondria. J Bioenerg Biomembr. 41: 361–367.

Burki F, et al. 2007. Phylogenomics reshuffles the eukaryotic supergroups. PLoS One 2:e790.

Burki F, et al. 2009. Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. Genome Biol Evol. 1: 231–238.

Burki F, Shalchian-Tabrizi K, Pawlowski J. 2008. Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. Biol Lett. 4:366–369.
Cavalier-Smith T. 2006. Origin of mitochondria by intracellular enslavement of a photosynthetic purple bacterium. Proc Biol Sci. 273:1943–1952.
Chan NC, Lithgow T. 2008. The peripheral membrane subunits of the SAM complex function codependently in mitochondrial outer membrane biogenesis. Mol Biol Cell. 19:126–136.
Colombini M. 2004. VDAC: the channel at the interface between mitochondria and the cytosol. Mol Cell Biochem. 256–257:107–115.
Dacks JB, Walker G, Field MC. 2008. Implications of the new eukaryotic systematics for parasitologists. Parasitol Int. 57:97–104.
Dacks JB, et al. 2002. Analyses of RNA Polymerase II genes from free-living protists: phylogeny, long branch attraction, and the eukaryotic big bang. Mol Biol Evol. 19:830–840.
Das S, Nikoladis N, Klein J, Masatoshi N. 2008. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. Proc Natl Acad Sci U S A. 28:16647–16652.
De Pinto V, et al. 2010. Characterization of human VDAC isoforms: a peculiar function for VDAC3? Biochim Biophys Acta. 1797: 1268–1275.
Dolezal P, Likic V, Tachezy J, Lithgow T. 2006. Evolution of the molecular machines for protein import into mitochondria. Science 313: 314–318.
Dolezal P, et al. 2010. The essentials of protein import in the degenerate mitochondrion of Entamoeba histolytica. PLoS Pathog. 6:e1000812.
Eddy SR. 1998. Profile hidden Markov models. Bioinformatics 14: 755–763.
Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference. Genome Inform. 23:205–211.
Eichinger L, Noegel AA. 2003. Crawling into a new era-the Dictyostelium genome project. EMBO J. 22:1941–1946.
Eichinger L, et al. 2005. The genome of the social amoeba Dictyostelium discoideum. Nature 435:43–57.
Finn RD, et al. 2010. The Pfam protein families database. Nucleic Acids Res. 38:D211–D222.
Gentle I, et al. 2004. The Omp85 family of proteins is essential for outer membrane biogenesis in mitochondria and bacteria. J Cell Biol. 164:19–24.
Gonçalves RP, et al. 2007. Supramolecular assembly of VDAC in native mitochondrial outer membranes. J Mol Biol. 369:413–418.
Gray MW, Burger G, Lang BF. 1999. Mitochondrial evolution. Science 283:1476–1481.
Gromiha MM, Ahmad S, Suwa M. 2005. TMBETA-NET: discrimination and prediction of membrane spanning beta-strands in outer membrane proteins. Nucleic Acids Res. 33:W164–W167.
Hackett JD, et al. 2007. Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. Mol Biol Evol. 24:1702–1713.
Hampl V, et al. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups." Proc Natl Acad Sci U S A. 106:3859–3864.
Hiller S, et al. 2008. Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. Science 321:1206–1210.
Hoogenboom BW, Suda K, Engel A, Fotiadis D. 2007. The supramolecular assemblies of voltage-dependent anion channels in the native membrane. J Mol Biol. 370:246–255.
Jarmuszkiewicz W, Sluse FE, Hryniewiecka L, Sluse-Goffart CM. 2002. Interactions between the cytochrome pathway and the alternative oxidase in isolated Acanthamoeba castellanii mitochondria. J Bioenerg Biomembr. 34:31–40.
Jarmuszkiewicz W, Wagner AM, Wagner MJ, Hryniewiecka L. 1997. Immunological identification of the alternative oxidase of Acanthamoeba castellanii mitochondria. FEBS Lett. 411:110–114.
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
Keeling PJ, et al. 2005. The tree of eukaryotes. Trends Ecol Evol. 20: 670–676.
Kicinska A, et al. 2007. ATP-sensitive potassium channel in mitochondria of the eukaryotic microorganism Acanthamoeba castellanii. J Biol Chem. 282:17433–17441.
Kinoshita JY, Mihara K, Oka T. 2007. Identification and characterization of a new tom40 isoform, a central component of mitochondrial outer membrane translocase. J Biochem. 141:897–906.
Kutik S, Stroud DA, Wiedemann N, Pfanner N. 2009. Evolution of mitochondrial protein biogenesis. Biochim Biophys Acta. 1790: 409–415.
Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23: 127–128.
Lister R, Hulett JM, Lithgow T, Whelan J. 2005. Protein import into mitochondria: origins and functions today (review). Mol Membr Biol. 22:87–100.
Maćasev D, et al. 2004. Tom22′, an 8-kDa trans-site receptor in plants and protozoans, is a conserved feature of the TOM complex that appeared early in the evolution of eukaryotes. Mol Biol Evol. 21: 1557–1564.
Mannella CA, Kinnally KW. 2008. Reflections on VDAC as a voltage-gated channel and a mitochondrial regulator. J Bioenerg Biomembr. 40:149–155.
Mayer A, Lill R, Neupert W. 1993. Translocation and insertion of precursor proteins into isolated outer membranes of mitochondria. J Cell Biol. 121:1233–1243.
Minge MA, et al. 2009. Evolutionary position of breviate amoebae and the primary eukaryote divergence. Proc Biol Sci. 276:597–604.
Mullis KB, Ferre F, Gibbs RA. 1994. The polymerase chain reaction. Boston: Birkhauser.
Natt NK, Kaur H, Raghava GP. 2004. Prediction of transmembrane regions of beta-barrel proteins using ANN- and SVM-based methods. Proteins 56:11–18.
Neupert W, Herrmann JM. 2007. Translocation of proteins into mitochondria. Annu Rev Biochem. 76:723–749.
Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol. 302:205–217.
O'Brien EA, et al. 2007. TBestDB: a taxonomically broad database of expressed sequence tags (ESTs). Nucleic Acids Res. 35: D445–D451.
Parfrey LW, et al. 2006. Evaluating support for the current classification of eukaryotic diversity. PLoS Genet. 2(12):e220.
Paschen SA, Neupert W, Rapaport D. 2005. Biogenesis of beta-barrel membrane proteins of mitochondria. Trends Biochem Sci. 30:575–582.
Paschen SA, et al. 2003. Evolutionary conservation of biogenesis of beta-barrel membrane proteins. Nature 426:862–866.
Perry AJ, et al. 2008. Structure, topology and function of the translocase of the outer membrane of mitochondria. Plant Physiol Biochem. 46:265–274.

Poirot O, O'Toole E, Notredame C. 2003. Tcoffee@igs: a web server for computing, evaluating and combining multiple sequence alignments. Nucleic Acids Res. 31:3503–3506.
Quevillon E, et al. 2005. InterProScan: protein domains identifier. Nucleic Acids Res. 33:W116–W120.
Rapaport D. 2005. How does the TOM complex mediate insertion of precursor proteins into the mitochondrial outer membrane? J Cell Biol. 171:419–423.
Rostovtseva TK, Bezrukov SM. 2008. VDAC regulation: role of cytosolic proteins and mitochondrial lipids. J Bioenerg Biomembr. 40:163–170.
Russell AG, Charette JM, Spencer DF, Gray MW. 2006. An early evolutionary origin for the minor spliceosome. Nature 443:863–866.
Ryan MT. 2004. Chaperones: inserting beta barrels into membranes. Curr Biol. 14:R207–R209.
Saccone C, et al. 2003. Molecular clock and gene function. J Mol Evol. 57(Suppl 1):S277–S285.
Sampson MJ, Lovell RS, Davison DB, Craigen WJ. 1996. A novel mouse mitochondrial voltage-dependent anion channel gene localizes to chromosome 8. Genomics 36:192–196.
Schulz GE. 2000. beta-Barrel membrane proteins. Curr Opin Struct Biol. 10:443–447.
Shoshan-Barmatz V, Israelson A, Brdiczka D, Sheu SS. 2006. The voltage-dependent anion channel (VDAC): function in intracellular signalling, cell life and cell death. Curr Pharm Des. 12:2249–2270.
Shoshan-Barmatz V, et al. 2010. VDAC, a multi-functional mitochondrial protein regulating cell life and death. Mol Aspects Med. 31: 227–285.
Shutt TE, Gray MW. 2006. Homologs of mitochondrial transcription factor B, sparsely distributed within the eukaryotic radiation, are likely derived from the dimethyladenosine methyltransferase of the mitochondrial endosymbiont. Mol Biol Evol. 23:1169–1179.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.
Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol. 57:758–771.
Thornton N, et al. 2010. Two modular forms of the mitochondrial sorting and assembly machinery are involved in biogenesis of alpha-helical outer membrane proteins. J Mol Biol. 396:540–549.
Tovar J, Fischer A, Clark CG. 1999. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite Entamoeba histolytica. Mol Microbiol. 32:1013–1021.
Troll H, et al. 1992. Purification, functional characterization, and cDNA sequencing of mitochondrial porin from Dictyostelium discoideum. J Biol Chem. 267:21072–21079.
Ujwal R, et al. 2008. The crystal structure of mouse VDAC1 at 2.3 A resolution reveals mechanistic insights into metabolite gating. Proc Natl Acad Sci U S A. 105:17742–17747.
Visvesvara GS. 2010. Amebic meningoencephalitides and keratitis: challenges in diagnosis and treatment. Curr Opin Infect Dis. 23:590–594.
Waller RF, et al. 2009. Evidence of a reduced and modified mitochondrial protein import apparatus in microsporidian mitosomes. Eukaryot Cell. 8:19–26.
Walther DM, Rapaport D, Tommassen J. 2009. Biogenesis of beta-barrel membrane proteins in bacteria and eukaryotes: evolutionary conservation and divergence. Cell Mol Life Sci. 66:2789–2804.
Wandrey M, Trevaskis B, Brewin N, Udvardi MK. 2004. Molecular and cell biology of a family of voltage-dependent anion channel porins in Lotus japonicus. Plant Physiol. 134:182–193.
Waterhouse AM, et al. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189–1191.
Watkins RF, Gray MW. 2006. The frequency of eubacterium-to-eukaryote lateral gene transfers shows significant cross-taxa variation within amoebozoa. J Mol Evol. 63:801–814.
Williams JG, Noegel AA, Eichinger L. 2005. Manifestations of multicellularity: Dictyostelium reports in. Trends Genet. 21: 392–398.
Wojtkowska M, et al. 2005. An inception report on the TOM complex of the Amoeba Acanthamoeba castellanii, a simple model protozoan in mitochondria studies. J Bioenerg Biomembr. 37:261–268.
Yamano K, Tanaka-Yamano S, Endo T. 2010. Mdm10 as a dynamic constituent of the TOB/SAM complex directs coordinated assembly of Tom40. EMBO Rep. 11:187–193.
Young MJ, Bay DC, Hausner G, Court DA. 2007. The evolutionary history of mitochondrial porins. BMC Evol Biol. 7:31.
Zeth K. 2010. Structure and evolution of mitochondrial outer membrane proteins of beta-barrel topology. Biochim Biophys Acta. 1797: 1292–1299.
Zeth K, Thein M. 2010. Porins in prokaryotes and eukaryotes: common themes and variations. Biochem J. 431:13–22.

Associate editor: Hidemi Watanabe