ELSEVIER

Biochimica et Biophysica Acta 1506 (2001) 147–162

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.bba-direct.com

Review

Comparative genomics and bioenergetics

Jose Castresana *

European Molecular Biology Laboratory (EMBL), Biocomputing Unit, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

Received 8 June 2001; received in revised form 10 October 2001; accepted 19 October 2001

Abstract

Bacterial and archaeal complete genome sequences have been obtained from a wide range of evolutionary lines, which allows some general conclusions about the phylogenetic distribution and evolution of bioenergetic pathways to be drawn. In particular, I searched in the complete genomes for key enzymes involved in aerobic and anaerobic respiratory pathways and in photosynthesis, and mapped them into an rRNA tree of sequenced species. The phylogenetic distribution of these enzymes is very irregular, and clearly shows the diverse strategies of energy conservation used by prokaryotes. In addition, a thorough phylogenetic analysis of other bioenergetic protein families of wide distribution reveals a complex evolutionary history for the respective genes. A parsimonious explanation for these complex phylogenetic patterns and for the irregular distribution of metabolic pathways is that the last common ancestor of Bacteria and Archaea contained several members of every gene family as a consequence of previous gene or genome duplications, while different patterns of gene loss occurred during the evolution of every gene family. This would imply that the last universal ancestor was a bioenergetically sophisticated organism. Finally, important steps that occurred during the evolution of energetic machineries, such as the early evolution of aerobic respiration and the acquisition of eukaryotic mitochondria from a proteobacterium ancestor, are supported by the analysis of the complete genome sequences.

Keywords: Genomics; Aerobic respiration; Anaerobic respiration; Photosynthesis; Last universal ancestor; Molecular evolution

1. Introduction

The number of completely sequenced genomes has been rapidly increasing since the first complete genome, that of the bacterium *Haemophilus influenza*, was obtained [1]. The range of species sequenced to date, although pathogens and endoparasites in a big proportion, cover a large number of evolutionary lines within Archaea and Bacteria [2–40]. In these microbial genomes, approximately 50–70% (58% on average) of the number of predicted genes show homology with a gene of known function. Comparative analysis of these genes is helping to better understand which gene products work in metabolic pathways of different bacterial species and how gene families are distributed among different evolutionary lines [41–44]. In addition, the existence of a large fraction of unknown genes implies that classical genetic and biochemical studies as well as genome-wide functional analysis [45,46] are going to play an important role before the functions of all novel genes uncovered by genome sequencing projects can be identified.

Archaea and Bacteria can be found in almost all possible niches on Earth thanks to the enormous

* Fax: +49-6221-387-306.
E-mail address: jose.castresana@embl-heidelberg.de (J. Castresana).

0005-2728/01/$ – see front matter
PII: S0005-2728(01)00227-4

variety of energy sources that different species are able to exploit. Bioenergetic routes and particularly respiratory mechanisms are known to produce an important impact on the global geochemical cycles [47–50]. In addition, the evolutionary analysis of these pathways may help to better interpret the fossil and biogeochemical record [51]. In this review, I show the phylogenetic distribution of several important respiratory pathways in prokaryotes according to the presence or absence of key metabolic enzymes in the complete genome sequences. Furthermore, I construct trees of protein families with enough phylogenetic information and discuss possible evolutionary scenarios that can explain the complex patterns that arise from these trees. Finally, I show how genome sequence data are helping to better understand some important aspects and revealing new details of the evolution of bioenergetic systems.

2. Databases and methods

The count of the number of genes and different enzymatic reactions involved in energy metabolism in every species was taken, after simple text parsing, from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.ad.jp/kegg/) [52], that contained this information for 43 prokaryotic genomes of 37 different species (in the April 2001 release). For the genes, all subunits and all members of a multigene family were added to the final count in order to reflect the total number of different gene products involved in bioenergetic pathways. For the enzymatic reactions, only the number of different EC numbers was considered in order to reflect the real diversity of bioenergetic mechanisms.

Alignments of gene families involved in energy production and conversion were retrieved from the COG (clusters of orthologous genes) database (http://www.ncbi.nlm.nih.gov/COG/) [41,53], that included information for 34 complete genomes (including yeast).

The presence of genes encoding key enzymes involved in important bioenergetic pathways was confirmed by BLAST searches [54] against the set of genes in the complete genomes available in the KEGG database.

Small rRNA sequences of species with completely sequenced genomes were retrieved already aligned from the Ribosomal Database Project (RDP) (http://www.cme.msu.edu/RDP/html/) [55]. The sequences of *Aquifex aeolicus*, *Thermoplasma volcanium*, *Bacillus halodurans*, *Chlamydia pneumoniae*, *Buchnera* sp. and *Halobacterium* sp., not available in this database, were extracted from the complete genome sequences, aligned with ClustalW [56] to the RDP alignment, and slightly adjusted to follow the best possible alignment with the closest species in the RDP set.

Conserved blocks of the alignments of COG families and the small rRNA data set were selected with the Gblocks 0.90 program (http://www.embl-heidelberg.de/~castresa/Gblocks/) [57], that extracts alignment positions that can be reliably used in phylogenetic analysis, thus making fully automated, large-scale phylogenetic analysis possible. For the COG families, Gblocks was used with default parameters. For the small rRNA alignment, the 'Minimum Length Of A Block' parameter was set to 5 in order to select many short conserved blocks that are present in rRNA alignments.

Phylogenetic analysis of the COG families was performed by maximum likelihood using the MOLPHY 2.3 package [58] with the Dayhoff model of amino acid substitution. Tree searches were done by local rearrangement from an initial neighbor-joining tree. Trees were rooted by midpoint except in a few cases where it was possible to place the root between Archaea and Bacteria. Trees of the small rRNA alignment were calculated by maximum likelihood with PAUP 4.0b8 [59], using a branch-swapping search strategy starting from a neighbor-joining tree. The HKY model of nucleotide substitution with rate heterogeneity (six rate categories and an assumed proportion of invariable sites) was used. Parameters of the model were estimated from the initial neighbor-joining tree.

3. The tree of life

Fig. 1 shows a phylogenetic tree derived from the small rRNA of 37 bacterial and archaeal species for which the complete genome has been sequenced. Eukaryotic species, the members of the third major division of life [60], were not included in this tree since

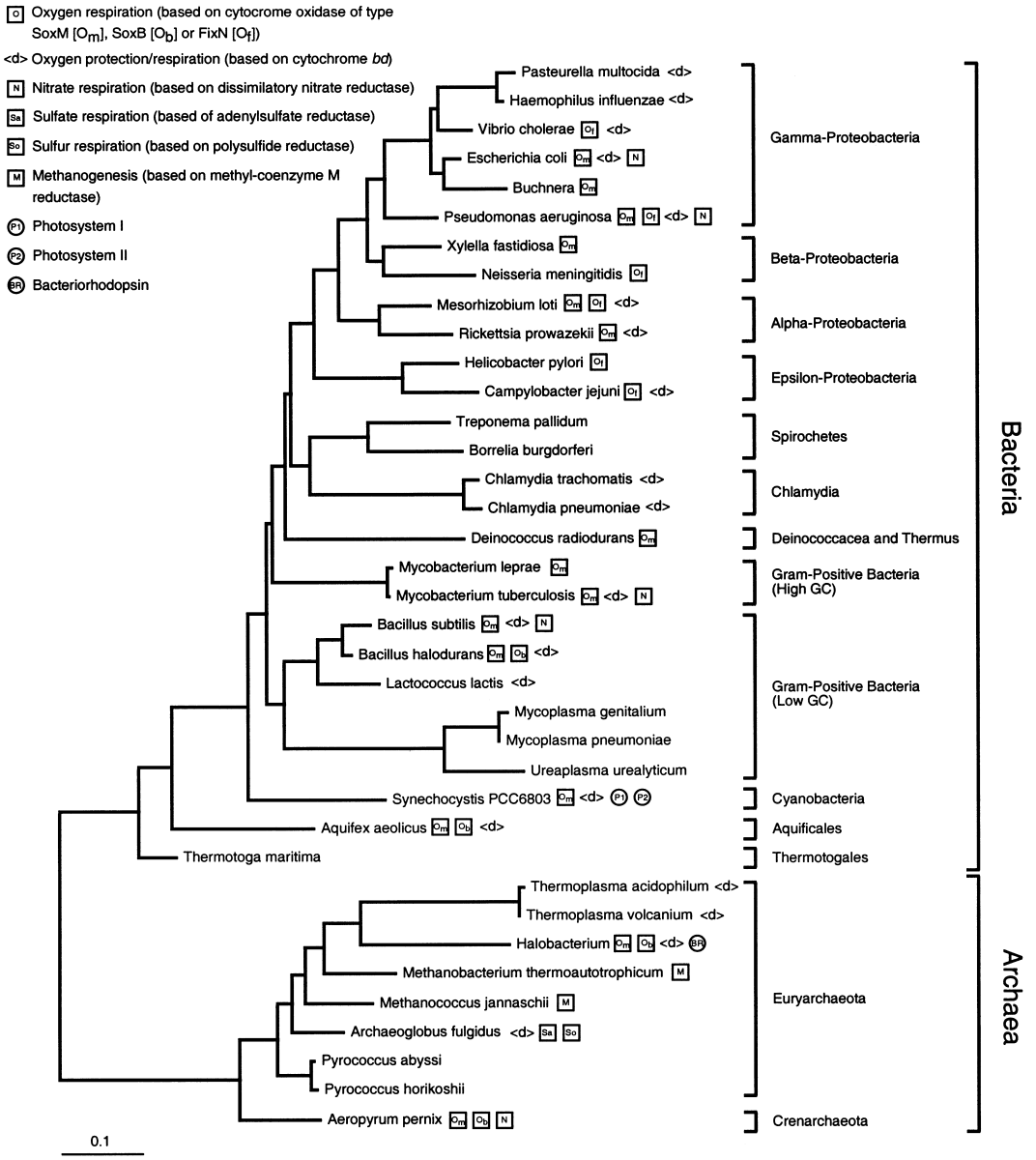


Fig. 1. A phylogenetic tree of bacterial and archaeal species whose genomes have been completely sequenced, based on the small rRNA sequence. The scale bar represents 0.1 nucleotide substitution per position. Bioenergetic respiratory pathways (as deduced from the presence of key enzymes) and photosynthetic pathways are mapped into the tree. Even if there are several members of each family or subfamily of key enzymes, only one symbol is shown. Most species display in addition several ways of fermentative and glycolytic reactions that are not shown.

they are, from a bioenergetic perspective, proteobacteria or cyanobacteria, due to the bacterial origin of mitochondria and chloroplasts [61–63]. The small rRNA gene is one of the few genes that has not experienced extensive gene duplications and losses, and it is conserved enough to produce reliable trees of the most divergent species of the tree of life [55,64–67]. Phylogenetic reconstruction methods such as the maximum likelihood method used for this tree do not give information about the root of the tree, i.e. about the position of the deepest split in the tree, and it is necessary to use some external method or additional knowledge to define such root. Since eukaryotic species were not included in

this tree, there is little doubt that the root should be placed between Archaea and Bacteria, two groups that differ in many molecular features. If eukaryotes were also considered, the place of the root would be more problematic. There were earlier data that supported a root separating Bacteria from an eukaryotes/Archaea group [68–70], although more recent analyses are challenging this view [71–73].

The species sequenced so far cover a wide range of prokaryotic evolutionary lines, including eight bacterial major divisions (proteobacteria, spirochetes, Deinococcaceae/Thermus group, Chlamydia, Gram-positive Bacteria, cyanobacteria, Aquificales and Thermotogales) and the two major groups in which Archaea are divided: Euryarchaeota and Crenarchaeota [74,75]. The overall structure of the tree in Fig. 1 is very similar to previous phylogenetic analysis based on rRNA that covered a larger number of species [75]. A noteworthy exception is that in the present tree (based on 1028 conserved positions selected with the Gblocks program, and with variable rates in the maximum likelihood evolutionary model) the earliest branching Bacteria is *Thermotoga*, whereas other analyses placed *Aquifex* branching earlier [75–77]. Even with a clean alignment and a sophisticated phylogenetic method, it is difficulty to correctly place these species in the tree due to the extremely low rate of evolution of these early-branching Bacteria (as it can be deduced from the short branches in the tree), which makes them prone to some tree reconstruction artifacts. In addition, some relationships among Archaea were unstable with varying sets of species. An exhaustive phylogenetic search of only archaeal sequences, and thus with more available conserved positions, shows that *Archaeoglobus fulgidus* groups with *Methanococcus jannaschii* rather than lying outside the methanogens group (not shown). Despite these phylogenetic uncertainties, the basic structure of the tree shown in Fig. 1 allows to understand the relationships and diversity of the species studied, and to map the main bioenergetics reactions into it.

4. Diversity of energy metabolism in prokaryotes from a genomic perspective

The main bioenergetic activities of sequenced species are shown in Fig. 1. Several enzymes involved in key respiratory reactions have been specifically checked with BLAST searches [54] in the KEGG database [52] and mapped, together with photosynthetic pathways, into the phylogenetic tree. All organisms in the tree display in addition several ways of fermentative and glycolytic reactions that have not been analyzed and are not shown in the figure.

After fermentative reactions, Fig. 1 shows that the bioenergetic activity that is most widely distributed in these species is aerobic metabolism. The key enzyme of this respiratory activity is cytochrome oxidase, that is known to constitute a large superfamily of homologous sequences [78–81]. The phylogenetic and evolutionary analysis of this family indicated that there are three basic types of cytochrome oxidase from the sequence point of view [82,83]: the SoxM and SoxB types (named after their genes in *Sulfolobus acidocaldarius* [84,85]), and the microaerobic FixN type (named after the genes in the nitrogen-fixing bacterium *Bradyrhizobium japonicum* [86]), which constitutes the most divergent group of the family. Most functional studies have been carried out with the SoxM type, which is also the form found in mitochondria. Whereas the SoxM and SoxB types work under normal aerobic conditions, the FixN oxidases (or cbb_3 type) are involved in respiration with very low oxygen pressure [86]. Many organisms use multiple oxidases belonging to the same or to different subfamilies for their aerobic metabolism, although the reasons for this apparent redundancy, whether flexibility to face different aerobic conditions or other reasons, are not always clear [87]. Interestingly, many pathogens (*Neisseria*, *Vibrio*, *Campylobacter*, *Helicobacter*) depend exclusively on high affinity oxidases for their microaerobic metabolism.

Cytochrome *bd* is a terminal oxidase with high affinity for oxygen, completely unrelated to the cytochrome oxidase superfamily. It may have an important function in coping with oxidative stress, although it has been shown that it is also able to create an electrochemical membrane gradient that is available for energetic requirements [87–89]. It is widely distributed among Archaea and Bacteria, and in a large proportion of cases it is present together with one or more cytochrome oxidases (Fig. 1). The function and expression of cytochrome *bd*

may differ among bacterial species, but the existence of multiple oxidases, that sometimes remained undetected, complicates the precise knowledge of the function of this enzyme. This can be alleviated with the information of the complete set of oxidases obtained from the genome sequences [87]. The cytochrome *bd* protein family consists of at least two main groups [90] that are the likely result of an ancient gene duplication (phylogenetic analysis not shown), but these two subfamilies have not been differentiated in Fig. 1.

Denitrification is an anaerobic respiratory pathway known to have also a wide phylogenetic distribution [91–94]. In contrast to aerobic respiration, where oxygen is the only terminal acceptor of electrons, in denitrification nitrate (NO_3^-) is sequentially reduced to dinitrogen (N_2) via nitrite (NO_2^-), nitric oxide (NO) and nitrous oxide (N_2O). Dissimilatory membrane-bound nitrate reductase is involved in the first step, reducing nitrate to nitrite, in a reaction coupled to energy conservation, and it has been found in several bacterial lineages and in one archaeum (Fig. 1) (or probably two [94]). However, the other enzymes of the denitrification process, nitrite reductase (of cytochrome cd_1 type), nitric oxide reductase and nitrous oxide reductase, whose sequences and biochemical activities are well known in *Pseudomonas aeruginosa* [92], are not found in the genomes of the other species that possess the first denitrification enzyme. A homologue of the copper-containing nitrite reductase (different from the cytochrome cd_1 type) is present in *Neisseria meningitidis*; in addition, a homologous enzyme distantly related to the nitric oxide reductase, qNOR, is present in this species as well as in *Synechocystis* sp., but these organisms are not denitrifying and therefore the function of the two latter enzymes may not be respiratory [95].

Other metabolic pathways have a more restricted distribution. Sulfate and sulfur respiration are confined, in the current set of sequenced species, to *A. fulgidus*, at least according to the distribution of adenylsulfate reductase [96,97] and polysulfide reductase [8,98,99], two complexes known to be involved, respectively, in these two types of anaerobic respiration. In addition, *A. fulgidus* is the only species of this set that contains the sulfite reductase [100,101]. These three enzymes are also known in several bacterial species [97–99,101,102] but their genomes have not been sequenced so far and they do not appear in the tree. Although *Thermoplasma* is also able to gain energy by sulfur respiration, no homologous enzymes to the ones known in *A. fulgidus* have been found in the two *Thermoplasma* species sequenced, and it has been proposed that other enzymes involved in sulfur metabolism could take this role [31].

Genes involved in all important steps of methanogenesis are found in the two methanogens sequenced so far: *M. jannaschii* and *Methanobacterium thermoautotrophicum* (Fig. 1). *A. fulgidus* possesses many genes involved in methanogenesis although it lacks the key enzyme methyl-coenzyme M reductase [103,104], therefore eliminating the possibility of methane production by the conventional pathway [8]. Finally, oxygenic photosynthesis, with both photosystems I and II [105,106], and bacteriorhodopsin-based photosynthesis are obviously only found in the organisms known to display these metabolic activities: cyanobacteria and halobacteria, respectively.

The patchy distribution of key metabolic enzymes shown in Fig. 1 is consistent with the diverse spectrum of bioenergetic strategies used by prokaryotes [47,107,108] and makes any prediction about the metabolism of certain species based on a close relative difficult. Gene loss seems to be the most important evolutionary force behind this irregular distribution [109,110] (see Section 5), even affecting crucial metabolic enzymes such as the ones analyzed here. The current set of sequenced species includes a large proportion of parasites, endosymbionts and pathogens at some other stage of their life cycles (most of the proteobacteria, spirochetes, chlamydias, mycobacteria and mycoplasmas – including *Ureoplasma*), that during their adaptation to this lifestyle from the free-living ancestors were very likely to loose many important metabolic routes [7,12,38,111,112]. However, even in free-living bacteria such as *Bacillus* it seems that, for example, *B. halodurans* lost the nitrate reductase while *B. subtilis* lost the SoxB-type cytochrome oxidase, all in a relatively short evolutionary period (Fig. 1).

Another way of viewing the bioenergetic diversity in these species and of understanding which ones have really experienced a reductive evolution in the bioenergetic machinery is to count the total number of genes, as well as the different enzymatic activities

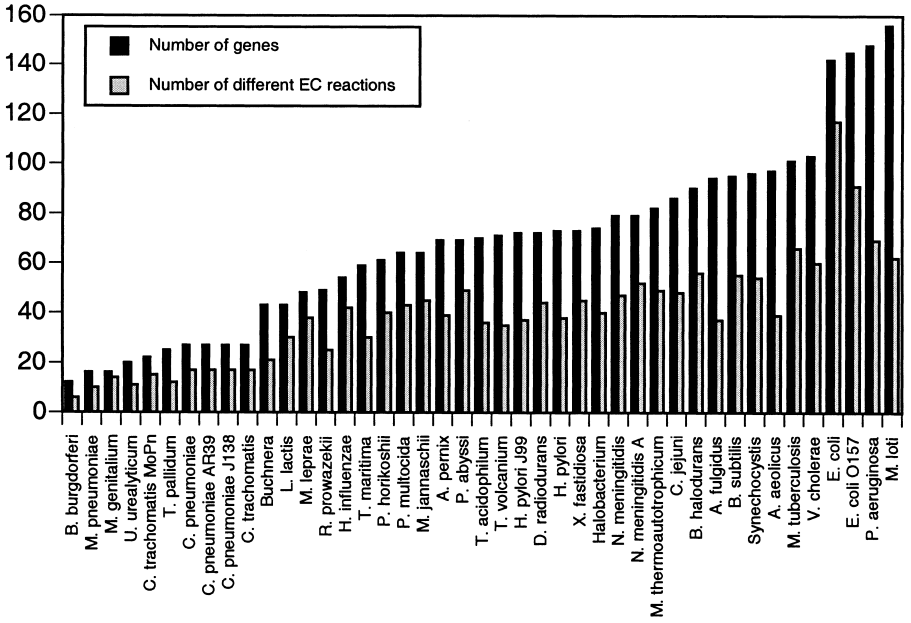


Fig. 2. Total number of genes and number of different enzymatic reactions (based on the EC number) assigned to energetic metabolism in bacterial and archaeal species with complete genomes sequenced, calculated from the KEGG database. Species are sorted according to the number of genes.

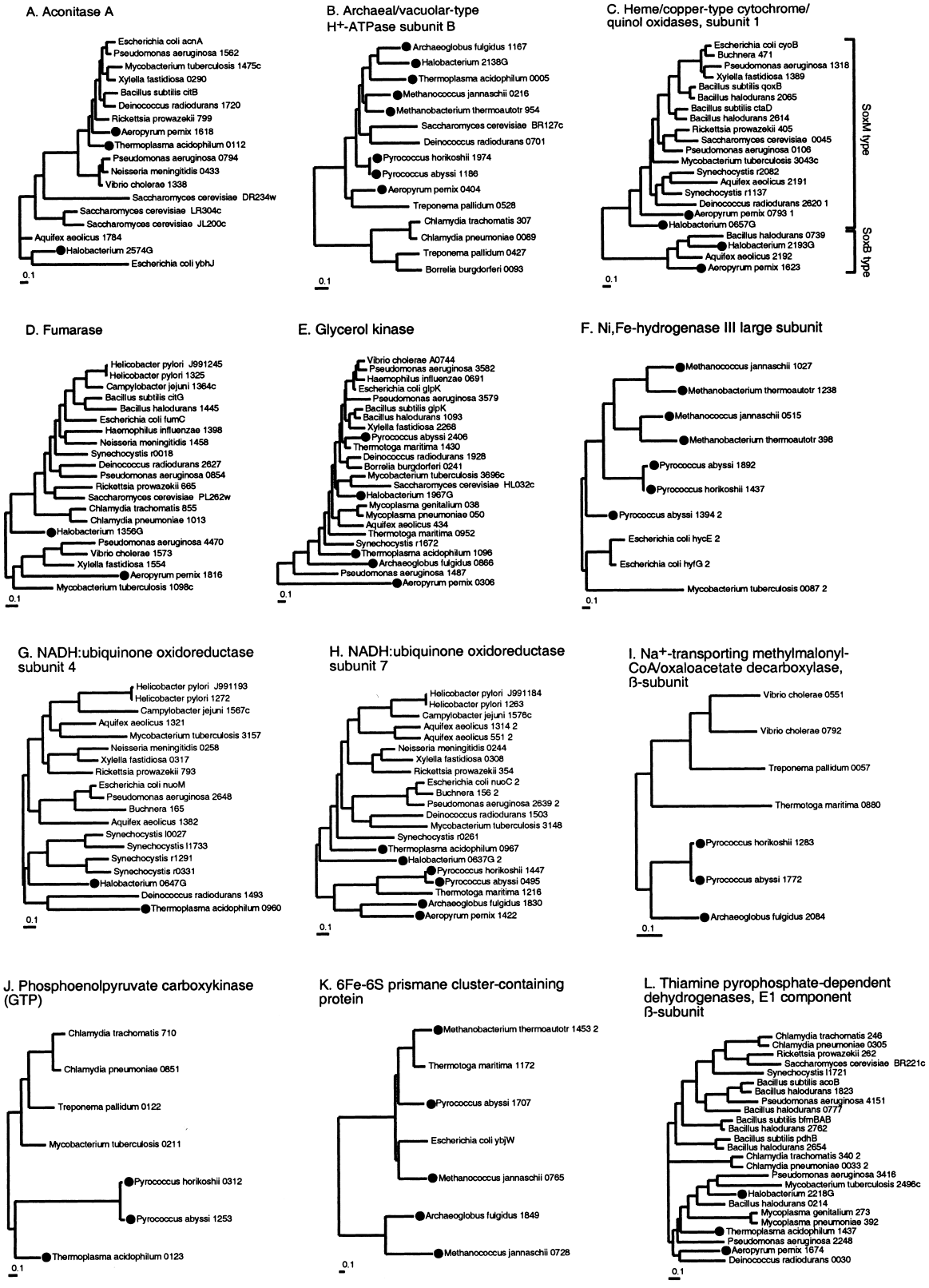
(EC numbers), assigned to energy conversion jobs (Fig. 2). These numbers roughly reflect the bioenergetic complexity of every species. According to the classification made by the KEGG database [52], some proteobacteria able of a free-living subsistence, such as *Escherichia coli*, *P. aeruginosa* or the nitrogen-fixing bacterium *Mesorhizobium loti*, dedicate around 150 genes to bioenergetics, while some parasites such as spirochetes (*Borrelia* and *Treponema*), chlamydias and mycoplasmas only employ around 20 genes or less for this activity. Despite the difficulties in assigning some enzymes to broad functional categories [113,114], the chart in Fig. 2 gives an approximate idea of the different number of protein families that can be found in bioenergetic reactions in these species. The picture of metabolic complexity obtained from the count of genes involved in energy production and conversion according to the COG database was very similar (not shown) despite some differences in the classification scheme used. The lack of phylogenetic grouping of the bioenergetically simplest organisms shows that strong reductive evolution of bioenergetic systems as a consequence of adaptation to a very specialized niche (such as the interior of a cell) has occurred often in evolution.

5. Phylogenetic trees of gene families involved in energy metabolism and evolutionary implications

Although key enzymes help to easily define the presence of bioenergetic routes in sequenced microorganisms, there are many other proteins working in these pathways whose analysis may improve our global overview of energetic metabolism. In the COG database there are 210 families (or clusters of orthologous genes) involved in energy production and conversion, and 141 of them are present in at least one archaeal and one bacterial species. Alignments with very few positions or those that are very

→

Fig. 3. Phylogenetic trees of protein families from the COG database with more than 200 reliably aligned positions extracted by the Gblocks program. Only trees of families present in at least two bacterial and two archaeal species are shown. The scale bar represents 0.1 amino acid substitution per position. Archaeal species are shown with a circle in front of the name. The numbers or letters following every species name are included to fit the name used by the COG database. In the cytochrome oxidase tree (C), the SoxM- and SoxB-type oxidase subfamilies are marked.


A. Aconitase A
Escherichia coli acnA
Pseudomonas aeruginosa 1562
Mycobacterium tuberculosis 1475c
Xylella fastidiosa 0290
Bacillus subtilis citB
Deinococcus radiodurans 1720
Rickettsia prowazekii 799
Aeropyrum pernix 1618
Thermoplasma acidophilum 0112
Pseudomonas aeruginosa 0794
Neisseria meningitidis 0433
Vibrio cholerae 1338
Saccharomyces cerevisiae DR234w
Saccharomyces cerevisiae LR304c
Saccharomyces cerevisiae JL200c
Aquifex aeolicus 1784
Halobacterium 2574G
Escherichia coli ybhJ
0.1
B. Archaeal/vacuolar-type H+-ATPase subunit B
Archaeoglobus fulgidus 1167
Halobacterium 2138G
Thermoplasma acidophilum 0005
Methanococcus jannaschii 0216
Methanobacterium thermoautotr 954
Saccharomyces cerevisiae BR127c
Deinococcus radiodurans 0701
Pyrococcus horikoshii 1974
Pyrococcus abyssi 1186
Aeropyrum pernix 0404
Treponema pallidum 0528
Chlamydia trachomatis 307
Chlamydia pneumoniae 0089
Treponema pallidum 0427
Borrelia burgdorferi 0093
0.1
C. Heme/copper-type cytochrome/ quinol oxidases, subunit 1
Escherichia coli cyoB
Buchnera 471
Pseudomonas aeruginosa 1318
Xylella fastidiosa 1389
Bacillus subtilis qoxB
Bacillus halodurans 2065
Bacillus subtilis ctaD
Bacillus halodurans 2614
Rickettsia prowazekii 405
Saccharomyces cerevisiae 0045
Pseudomonas aeruginosa 0106
Mycobacterium tuberculosis 3043c
Synechocystis r2082
Aquifex aeolicus 2191
Synechocystis r1137
Deinococcus radiodurans 2620 1
Aeropyrum pernix 0793 1
Halobacterium 0657G
Bacillus halodurans 0739
Halobacterium 2193G
Aquifex aeolicus 2192
Aeropyrum pernix 1623
SoxM type
SoxB type
0.1
D. Fumarase
Helicobacter pylori J991245
Helicobacter pylori 1325
Campylobacter jejuni 1364c
Bacillus subtilis citG
Bacillus halodurans 1445
Escherichia coli fumC
Haemophilus influenzae 1398
Neisseria meningitidis 1458
Synechocystis r0018
Deinococcus radiodurans 2627
Pseudomonas aeruginosa 0854
Rickettsia prowazekii 665
Saccharomyces cerevisiae PL262w
Chlamydia trachomatis 855
Chlamydia pneumoniae 1013
Halobacterium 1356G
Pseudomonas aeruginosa 4470
Vibrio cholerae 1573
Xylella fastidiosa 1554
Aeropyrum pernix 1816
Mycobacterium tuberculosis 1098c
0.1
E. Glycerol kinase
Vibrio cholerae A0744
Pseudomonas aeruginosa 3582
Haemophilus influenzae 0691
Escherichia coli glpK
Pseudomonas aeruginosa 3579
Bacillus subtilis glpK
Bacillus halodurans 1093
Xylella fastidiosa 2268
Pyrococcus abyssi 2406
Thermotoga maritima 1430
Deinococcus radiodurans 1928
Borrelia burgdorferi 0241
Mycobacterium tuberculosis 3696c
Saccharomyces cerevisiae HL032c
Halobacterium 1967G
Mycoplasma genitalium 038
Mycoplasma pneumoniae 050
Aquifex aeolicus 434
Thermotoga maritima 0952
Synechocystis r1672
Thermoplasma acidophilum 1096
Archaeoglobus fulgidus 0866
Pseudomonas aeruginosa 1487
Aeropyrum pernix 0306
0.1
F. Ni,Fe-hydrogenase III large subunit
Methanococcus jannaschii 1027
Methanobacterium thermoautotr 1238
Methanococcus jannaschii 0515
Methanobacterium thermoautotr 398
Pyrococcus abyssi 1892
Pyrococcus horikoshii 1437
Pyrococcus abyssi 1394 2
Escherichia coli hycE 2
Escherichia coli hyfG 2
Mycobacterium tuberculosis 0087 2
0.1
G. NADH:ubiquinone oxidoreductase subunit 4
Helicobacter pylori J991193
Helicobacter pylori 1272
Campylobacter jejuni 1567c
Aquifex aeolicus 1321
Mycobacterium tuberculosis 3157
Neisseria meningitidis 0258
Xylella fastidiosa 0317
Rickettsia prowazekii 793
Escherichia coli nuoM
Pseudomonas aeruginosa 2648
Buchnera 165
Aquifex aeolicus 1382
Synechocystis l0027
Synechocystis l1733
Synechocystis r1291
Synechocystis r0331
Halobacterium 0647G
Deinococcus radiodurans 1493
Thermoplasma acidophilum 0960
0.1
H. NADH:ubiquinone oxidoreductase subunit 7
Helicobacter pylori J991184
Helicobacter pylori 1263
Campylobacter jejuni 1576c
Aquifex aeolicus 1314 2
Aquifex aeolicus 551 2
Neisseria meningitidis 0244
Xylella fastidiosa 0308
Rickettsia prowazekii 354
Escherichia coli nuoC 2
Buchnera 156 2
Pseudomonas aeruginosa 2639 2
Deinococcus radiodurans 1503
Mycobacterium tuberculosis 3148
Synechocystis r0261
Thermoplasma acidophilum 0967
Halobacterium 0637G 2
Pyrococcus horikoshii 1447
Pyrococcus abyssi 0495
Thermotoga maritima 1216
Archaeoglobus fulgidus 1830
Aeropyrum pernix 1422
0.1
I. Na+-transporting methylmalonyl-CoA/oxaloacetate decarboxylase, β-subunit
Vibrio cholerae 0551
Vibrio cholerae 0792
Treponema pallidum 0057
Thermotoga maritima 0880
Pyrococcus horikoshii 1283
Pyrococcus abyssi 1772
Archaeoglobus fulgidus 2084
0.1
J. Phosphoenolpyruvate carboxykinase (GTP)
Chlamydia trachomatis 710
Chlamydia pneumoniae 0851
Treponema pallidum 0122
Mycobacterium tuberculosis 0211
Pyrococcus horikoshii 0312
Pyrococcus abyssi 1253
Thermoplasma acidophilum 0123
0.1
K. 6Fe-6S prismane cluster-containing protein
Methanobacterium thermoautotr 1453 2
Thermotoga maritima 1172
Pyrococcus abyssi 1707
Escherichia coli ybjW
Methanococcus jannaschii 0765
Archaeoglobus fulgidus 1849
Methanococcus jannaschii 0728
0.1
L. Thiamine pyrophosphate-dependent dehydrogenases, E1 component β-subunit
Chlamydia trachomatis 246
Chlamydia pneumoniae 0305
Rickettsia prowazekii 262
Saccharomyces cerevisiae BR221c
Synechocystis l1721
Bacillus subtilis acoB
Bacillus halodurans 1823
Pseudomonas aeruginosa 4151
Bacillus halodurans 0777
Bacillus subtilis bfmBAB
Bacillus halodurans 2762
Bacillus subtilis pdhB
Bacillus halodurans 2654
Chlamydia trachomatis 340 2
Chlamydia pneumoniae 0033 2
Pseudomonas aeruginosa 3416
Mycobacterium tuberculosis 2496c
Halobacterium 2218G
Bacillus halodurans 0214
Mycoplasma genitalium 273
Mycoplasma pneumoniae 392
Thermoplasma acidophilum 1437
Pseudomonas aeruginosa 2248
Aeropyrum pernix 1674
Deinococcus radiodurans 0030
0.1

divergent are not suitable for phylogenetic analysis [71,73]. The 141 protein alignments of wide distribution were checked with the Gblocks program [57] using default parameters to extract reliably aligned positions and thus suitable for phylogenetic analysis. According to the results, most of the alignments were quite obviously useless for phylogenetic analysis, with half of them having under 50 positions available for analysis. It is not clear when an alignment will contain enough phylogenetic information to produce a meaningful tree, so I arbitrarily chose those with more than 200 positions (although some of those with less positions will probably contain some valid phylogenetic information). This resulted in 16 alignments, representative of bioenergetic protein families, that were used to construct maximum likelihood phylogenetic trees. Fig. 3 shows 12 of them (those with at least two bacterial and two archaeal species). None of these gene trees reflects the species phylogeny as depicted in the small rRNA tree (Fig. 1) and they show, to a greater or lesser extent, a complicated evolutionary history. Even the most fundamental division, that between Archaea and Bacteria, is only directly reflected by three of the trees shown in the figure (Fig. 3F,I,J). Thus, the analysis performed here with representative proteins from bioenergetic pathways of wide distribution indicates that these proteins are not different from those involved in other metabolic activities, whose analysis also revealed many complex phylogenetic patterns [115].

Several reasons can account for the observed discrepancies in these trees. First, some species contain several members of the corresponding gene family, and therefore it is obvious that gene duplications and paralogy (see Fig. 4) are contributing in part to the complexity of the trees. For example, gene duplications occurred early in the evolution of cytochrome oxidases (Fig. 3C; the FixN type is considered a different group in the COG database) as well as later in the evolution of the SoxM-type oxidase [82,83]. Second, the large evolutionary periods that separate the sequences in all trees, that may involve changes in the mutation rates and modes of evolution in different parts of the trees, are surely contributing to tree reconstruction artifacts, that will particularly affect groups separated by short internal branches. Third, some sequences of related species are too separated in the tree (see, for example, the separation of *Halobacterium* from the two other archaeal sequences in the aconitase A tree; Fig. 3A) and in these cases the most satisfactory explanation is the occurrence of lateral gene transfer. Transfer of genetic material between different species must have occurred in many instances during bacterial evolution [21,116–120] although it has also been argued that this phenomenon has been over-emphasized in the literature, and that the first two mentioned possibilities are often overlooked [110,117, 121–124].

Another scenario that can explain in a more parsimonious way a major part of the complex patterns in the universal trees of bioenergetic proteins (and of other metabolic pathways) is that the last common ancestor of Archaea and Bacteria (which is also the last universal ancestor or LUCA, provided that there is a bacterial root in the tree of life) would have had several copies of every gene family [110]. It is not unlikely that this redundancy could have occurred as a consequence of partial or whole genome duplications [125–127] previous to the last universal ancestor. As different bacterial and archaeal lineages diverged and adapted to specialized niches, many species would have lost some copies of the multigene families, creating very complicate phylogenetic patterns and paralogies that would not reflect the real relationships of the species. Although, obviously, single gene duplications, lateral gene transfers and tree reconstruction artifacts must also have played an important role, the possibility of partial or whole genome duplications previous to the last universal an-

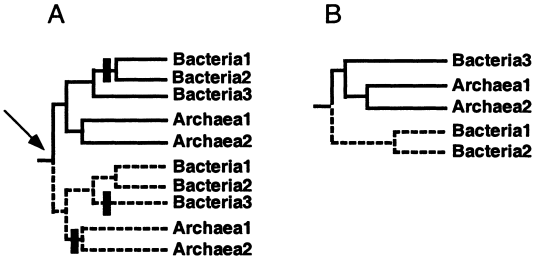


Fig. 4. (A) Schematic representation of a tree of all archaeal and bacterial sequences of a gene family resulting after a gene duplication (arrow) previous to the separation of Archaea and Bacteria. Thick lines indicate gene losses in the next tree. (B) Tree that compares the paralogous sequences resulting after the three independent gene losses in the previous phylogeny. The result is that Archaea and Bacteria are not longer separated in this tree.

cestor followed by lineage-specific gene losses can provide a unique explanation for a major part of the complex phylogenies found with many genes, as well as for the patchy distribution of bioenergetic pathways (see Section 4). This, in turn, would favor the idea of a complex and metabolically very versatile last universal ancestor, in contrast with Woese's progenote model that supposes that this organism was not a real cellular entity but rather a complex population of heterogeneous genetic entities that experienced genetic exchange at high frequency, a question that has been actively debated from different perspectives [71,99,110,128–132]. When more genomes become available it will be possible to spot more specifically in the trees of every protein family the exact points of gene duplications [133], at least for the proteins with enough phylogenetic information, and thus to establish the relative contribution of ancestral duplications and lateral gene transfer to the evolution of multigene families.

6. Genome data add support to the early evolution of aerobic respiration and the bacterial origin of eukaryotic mitochondria

A dogma of bioenergetic evolution has been for a very long time that photosynthesis was one of the earliest ways of generating energy for the cell, whereas aerobic metabolism appeared late in evolution, and in several independent lineages, after the release of oxygen by cyanobacteria. This reasoning was mainly based on the interpretation of certain geological and ancient fossil data. On the contrary, the phylogenetic analysis of cytochrome oxidase sequences indicated that aerobic organisms belonging to very diverse bacterial and archaeal groups used a homologous cytochrome oxidase and, therefore, that aerobic metabolism must have had a single and ancient origin [82,83,99,134], while the restricted distribution of oxygenic photosynthesis to cyanobacteria makes it more unlikely that this pathway was present in the last universal ancestor (the respiration-early hypothesis). Availability of oxygen for aerobic respiration would not have been a problem since there is plenty of evidence showing that a certain amount of free oxygen was present in the early Earth [135,136].

Genome data make clearer the wide distribution of cytochrome oxidase in Archaea and Bacteria (Fig. 1), and indicate again that this is an ancient enzyme that must have been present in the last universal ancestor. In addition, more organisms are being added to the list previously known [99,137] with both the SoxM- and SoxB-type oxidases: *Halobacterium* sp., *B. halodurans*, *A. aeolicus* and *Aeropyrum pernix*. These two types of cytochrome oxidase are a product of a gene duplication previous to the last universal ancestor [82,83] (see also cytochrome oxidase tree in Fig. 3C) indicating that this organism possessed at least both oxidase types. Although archaeal sequences do not appear perfectly monophyletic in the tree, both in the SoxM and in the SoxB parts of the tree, it is likely that tree reconstruction artifacts affecting short internal branches that separate the archaeal sequences, or further gene duplications and losses, explain this discrepancy. Transfer of genes between Archaea and Bacteria can be considered extremely unlikely for integral membrane proteins with a complex enzymatic activity like the cytochrome oxidase, due to the very different lipidic composition and biophysical properties of archaeal and bacterial membranes. These different properties cause a big reduction in the proton pumping activity of archaeal cytochrome oxidases working in bacterial lipids [138] and in the activity of bacterial oxidases reconstituted in archaeal lipids [139]. The existence of two types of cytochrome oxidase in the last universal ancestor would also make it very unlikely that the putative ancestral function of this enzyme, that probably was nitric oxide (NO) reduction [95,140–142], changed to its actual function late in evolution, since this change would have involved a complex evolutionary invention in several independent lineages, and not only in one type, but in two different types of cytochrome oxidase. Thus, the evolution from an NO reductase function to the actual oxidase function most likely took place only once and, therefore, much before the last universal ancestor.

Genome comparisons point to an aerobic last universal ancestor, but we find today many anaerobic organisms that have lost this respiratory capacity (Fig. 1). Ancestors of the eukaryotic lineage were probably among those that lost aerobic respiration very early but, since they needed it for a complex cellular metabolism, they later recovered

this capacity from a bacterium that had not lost it, giving rise to the mitochondria. The origin of eukaryotic mitochondria from a proteobacterial ancestor is well established from the comparison of several types of molecular data [61–63,143]. The tight interaction between the ancestral eukaryote host and the bacterial endosymbiont led to the transfer of many bacterial/mitochondrial genes to the nucleus of the host while some proteins of the mitochondria were recruited from the set of nuclear, eukaryotic lineage genes. The sequence and analysis of the complete genomes of yeast and several proteobacteria, including *Rickettsia prowazekii* [12], which is the closest species sequenced so far to the bacterial ancestor that gave rise to the mitochondria, have allowed to trace with great detail the evolutionary history of all proteins of the eukaryotic mitochondria. Surprisingly, half of the approximately 400 mitochondrial proteins in yeast have no discernible bacterial homologue and only 1/10 of them have their origins in the ancestral proteobacterium [12,144,145]. With respect to the main electron transfer complexes of the respiratory chain, the core subunits can be traced to bacterial genes but many of the other accessory and assembly subunits were adopted from the eukaryotic genome. In addition, most of the transporters of yeast mitochondria are of eukaryotic origin, and were targeted to the mitochondrial membranes, converting the mitochondria into an organelle with an ATP exporting function [144,145]. Thus, this thorough analysis of yeast and bacterial genomes shows that taming of mitochondria occurred in a much more expeditious way than previously thought.

7. Conclusions

Three important aspects of energetic metabolism can be learned from the comparison of microbial genomes. First, the distribution of key enzymes shows a patchy distribution of bioenergetic pathways in the tree of life (Fig. 1). Second, parasites and endosymbionts show an important reduction in their energetic machinery when they adapt to this highly specialized lifestyle (Fig. 2). Third, trees of protein families involved in energetic metabolism show very complex phylogenetic patterns (Fig. 3). These observations can be more parsimoniously explained if the last common ancestor of Archaea and Bacteria possessed many metabolic capacities and several members of every gene family, perhaps as a consequence of previous whole genome duplications, while lineage-specific gene losses created many paralogies in the gene trees of present-day organisms and an irregular distribution of bioenergetic pathways. Other phenomena such as lateral gene transfer may better explain the evolution of specific multigene families. Recognition of all possible types of evolutionary forces should guide whole genome comparisons in order to lead to a more detailed understanding of the molecular evolution and phylogenetic distribution of bioenergetic systems.

Acknowledgements

This paper is dedicated to the memory of Matti Saraste, who encouraged this work and realized the importance of embracing evolutionary thinking to better understand the cytochrome oxidase family.

References

[1] R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, B.A. Dougherty, J.M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J.D. Gocayne, J. Scott, R. Shirley, L.I. Liu, A. Glodek, J.M. Kelley, J.F. Weidman, C.A. Phillips, T. Spriggs, E. Hedblom, M.D. Cotton et al., Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, Science 269 (1995) 496–512.

[2] C.M. Fraser, J.D. Gocayne, O. White, M.D. Adams, R.A. Clayton, R.D. Fleischmann, C.J. Bult, A.R. Kerlavage, G. Sutton, J.M. Kelley, J.L. Fritchman, J.F. Weidman, K.V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T.R. Utterback, D.M. Saudek, C.A. Phillips, J.M. Merrick, J.-F. Tomb, B.A. Dougherty, K.F. Bott, P.-C. Hu, T.S. Lucier, S.N. Peterson, H.O. Smith, C.A. Hutchison, J.C. Venter, The minimal gene complement of *Mycoplasma genitalium*, Science 270 (1995) 397–403.

[3] C.J. Bult, O. White, G.J. Olsen, L. Zhou, R.D. Fleischmann, G.G. Sutton, J.A. Blake, L.M. FitzGerald, R.A. Clayton, J.D. Gocayne, A.R. Kerlavage, B.A. Dougherty, J.F. Tomb, M.D. Adams, C.I. Reich, R. Overbeek, E.F. Kirkness, K.G. Weinstock, J.M. Merrick, A. Glodek, J.L. Scott, N.S. Geoghagen, J.C. Venter, Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*, Science 273 (1996) 1058–1073.

[4] R. Himmelreich, H. Hilbert, H. Plagens, E. Pirkl, B.C. Li, R. Herrmann, Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*, Nucleic Acids Res. 24 (1996) 4420–4449.

[5] T. Kaneko, S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda, S. Tabata, Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions, DNA Res. 3 (1996) 109–136.

[6] F.R. Blattner, G. Plunkett III, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, Y. Shao, The complete genome sequence of *Escherichia coli* K-12, Science 277 (1997) 1453–1474.

[7] C.M. Fraser, S. Casjens, W.M. Huang, G.G. Sutton, R. Clayton, R. Lathigra, O. White, K.A. Ketchum, R. Dodson, E.K. Hickey, M. Gwinn, B. Dougherty, J.F. Tomb, R.D. Fleischmann, D. Richardson, J. Peterson, A.R. Kerlavage, J. Quackenbush, S. Salzberg, M. Hanson, R.V. Vugt, N. Palmer, M.D. Adams, J. Gocayne, J. Weidman et al., Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*, Nature 390 (1997) 580–586.

[8] H.-P. Klenk, R.A. Clayton, J.-F. Tomb, O. White, K.E. Nelson, K.A. Ketchum, R.J. Dodson, M. Gwinn, E.K. Hickey, J.D. Peterson, D.L. Richardson, A.R. Kerlavage, D.E. Graham, N.C. Kyrpides, R.D. Fleischmann, J. Quackenbush, N.H. Lee, G.G. Sutton, S. Gill, E.F. Kirkness, B.A. Dougherty, K. Mckenney, M.D. Adams, B. Loftus, S. Peterson et al., The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*, Nature 390 (1997) 364–370.

[9] F. Kunst, N. Ogasawara, I. Moszer, A.M. Albertini, G. Alloni, V. Azevedo, M.G. Bertero, P. Bessières, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S.C. Brignell, S. Bron, S. Brouillet, C.V. Bruschi, B. Caldwell, V. Capuano, N.M. Carter, S.-K. Choi, J.-J. Codani, I.F. Connerton, N.J. Cummings et al., The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*, Nature 390 (1997) 249–256.

[10] D.R. Smith, L.A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, J.N. Reeve, Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: functional analysis and comparative genomics, J. Bacteriol. 179 (1997) 7135–7155.

[11] J.-F. Tomb, O. White, A.R. Kerlavage, R.A. Clayton, G.G. Sutton, R.D. Fleischmann, K.A. Ketchum, H.P. Klenk, S. Gill, B.A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E.F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H.G. Khalak, A. Glodek, K. Mckenney, L.M. Fitzegerald, N. Lee, M.D. Adams, E.K. Hickey et al., The complete genome sequence of the gastric pathogen *Helicobacter pylori*, Nature 388 (1997) 539–547.

[12] S.G. Andersson, A. Zomorodipour, J.O. Andersson, T. Sicheritz-Ponten, U.C. Alsmark, R.M. Podowski, A.K. Naslund, A.S. Eriksson, H.H. Winkler, C.G. Kurland, The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria, Nature 396 (1998) 133–140.

[13] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh et al., Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.

[14] G. Deckert, P.V. Warren, T. Gaasterland, W.G. Young, A.L. Lenox, D.E. Graham, R. Overbeek, M.A. Snead, M. Keller, M. Aujay, R. Huber, R.A. Feldman, J.M. Short, G.J. Olsen, R.V. Swanson, The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*, Nature 392 (1998) 353–358.

[15] C.M. Fraser, S.J. Norris, G.M. Weinstock, O. White, G.G. Sutton, R. Dodson, M. Gwinn, E.K. Hickey, R. Clayton, K.A. Ketchum, E. Sodergren, J.M. Hardham, M.P. McLeod, S. Salzberg, J. Peterson, H. Khalak, D. Richardson, J.K. Howell, M. Chidambaram, T. Utterback, L. McDonald, P. Artiach, C. Bowman, M.D. Cotton, C. Fujii et al., Complete genome sequence of *Treponema pallidum*, the syphilis spirochete, Science 281 (1998) 375–388.

[16] Y. Kawarabayasi, M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, T. Yoshizawa et al., Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3, DNA Res. 5 (1998) 55–76.

[17] R.S. Stephens, S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R.L. Tatusov, Q. Zhao, E.V. Koonin, R.W. Davis, Genome sequence of an obligate intracellular pathogen of humans *Chlamydia trachomatis*, Science 282 (1998) 754–759.

[18] R.A. Alm, L.S. Ling, D.T. Moir, B.L. King, E.D. Brown, P.C. Doig, D.R. Smith, B. Noonan, B.C. Guild, B.L. deJonge, G. Carmel, P.J. Tummino, A. Caruso, M. Uria-Nickelsen, D.M. Mills, C. Ives, R. Gibson, D. Merberg, S.D. Mills, Q. Jiang, D.E. Taylor, G.F. Vovis, T.J. Trust, Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*, Nature 397 (1999) 176–180.

[19] S. Kalman, W. Mitchell, R. Marathe, C. Lammel, J. Fan,

R.W. Hyman, L. Olinger, J. Grimwood, R.W. Davis, R.S. Stephens, Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*, Nat. Genet. 21 (1999) 385–389.

[20] Y. Kawarabayasi, Y. Hino, H. Horikawa, S. Yamazaki, Y. Haikawa, K. Jin-no, M. Takahashi, M. Sekine, S. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, H. Nakazawa, M. Takamiya, S. Masuda, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki et al., Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1, DNA Res. 6 (1999) 83–101, 145–152.

[21] K.E. Nelson, R.A. Clayton, S.R. Gill, M.L. Gwinn, R.J. Dodson, D.H. Haft, E.K. Hickey, J.D. Peterson, W.C. Nelson, K.A. Ketchum, L. McDonald, T.R. Utterback, J.A. Malek, K.D. Linher, M.M. Garrett, A.M. Stewart, M.D. Cotton, M.S. Pratt, C.A. Phillips, D. Richardson, J. Heidelberg, G.G. Sutton, R.D. Fleischmann, J.A. Eisen, O. White, S.L. Salzberg, H.O. Smith, J.C. Venter, C.M. Fraser, Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*, Nature 399 (1999) 323–329.

[22] O. White, J.A. Eisen, J.F. Heidelberg, E.K. Hickey, J.D. Peterson, R.J. Dodson, D.H. Haft, M.L. Gwinn, W.C. Nelson, D.L. Richardson, K.S. Moffat, H. Qin, L. Jiang, W. Pamphile, M. Crosby, M. Shen, J.J. Vamathevan, P. Lam, L. McDonald, T. Utterback, C. Zalewski, K.S. Makarova, L. Aravind, M.J. Daly, K.W. Minton et al., Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1, Science 286 (1999) 1571–1577.

[23] J.I. Glass, E.J. Lefkowitz, J.S. Glass, C.R. Heiner, E.Y. Chen, G.H. Cassell, The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*, Nature 407 (2000) 757–762.

[24] J.F. Heidelberg, J.A. Eisen, W.C. Nelson, R.A. Clayton, M.L. Gwinn, R.J. Dodson, D.H. Haft, E.K. Hickey, J.D. Peterson, L. Umayam, S.R. Gill, K.E. Nelson, T.D. Read, H. Tettelin, D. Richardson, M.D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. Mcdonald, T. Utterback, R.D. Fleishmann, W.C. Nierman et al., DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*, Nature 406 (2000) 477–483.

[25] T. Kaneko, Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, A. Watanabe, K. Idesawa, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, Y. Mochizuki, S. Nakayama, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada, S. Tabata, Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*, DNA Res. 7 (2000) 331–338.

[26] T. Kawashima, N. Amano, H. Koike, S. Makino, S. Higuchi, Y. Kawashima-Ohya, K. Watanabe, M. Yamazaki, K. Kanehori, T. Kawamoto, T. Nunoshiba, Y. Yamamoto, H. Aramaki, K. Makino, M. Suzuki, Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*, Proc. Natl. Acad. Sci. USA 97 (2000) 14257–14262.

[27] W.V. Ng, S.P. Kennedy, G.G. Mahairas, B. Berquist, M. Pan, H.D. Shukla, S.R. Lasky, N.S. Baliga, V. Thorsson, J. Sbrogna, S. Swartzell, D. Weir, J. Hall, T.A. Dahl, R. Welti, Y.A. Goo, B. Leithauser, K. Keller, R. Cruz, M.J. Danson, D.W. Hough, D.G. Maddocks, P.E. Jablonski, M.P. Krebs, C.M. Angevine et al., Genome sequence of *Halobacterium* species NRC-1, Proc. Natl. Acad. Sci. USA 97 (2000) 12176–12181.

[28] J. Parkhill, M. Achtman, K.D. James, S.D. Bentley, C. Churcher, S.R. Klee, G. Morelli, D. Basham, D. Brown, T. Chillingworth, R.M. Davies, P. Davis, K. Devlin, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, S. Leather, S. Moule, K. Mungall, M.A. Quail, M.A. Rajandream, K.M. Rutherford, M. Simmonds, J. Skelton, S. Whitehead, B.G. Spratt, B.G. Barrell, Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491, Nature 404 (2000) 502–506.

[29] J. Parkhill, B.W. Wren, K. Mungall, J.M. Ketley, C. Churcher, D. Basham, T. Chillingworth, R.M. Davies, T. Feltwell, S. Holroyd, K. Jagels, A.V. Karlyshev, S. Moule, M.J. Pallen, C.W. Penn, M.A. Quail, M.A. Rajandream, K.M. Rutherford, A.H. vanVliet, S. Whitehead, B.G. Barrell, The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences, Nature 403 (2000) 665–668.

[30] T.D. Read, R.C. Brunham, C. Shen, S.R. Gill, J.F. Heidelberg, O. White, E.K. Hickey, J. Peterson, T. Utterback, K. Berry, S. Bass, K. Linher, J. Weidman, H. Khouri, B. Craven, C. Bowman, R. Dodson, M. Gwinn, W. Nelson, R. DeBoy, J. Kolonay, G. McClarty, S.L. Salzberg, J. Eisen, C.M. Fraser, Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39, Nucleic Acids Res. 28 (2000) 1397–1406.

[31] A. Ruepp, W. Graml, M.L. Santos-Martinez, K.K. Koretke, C. Volker, H.W. Mewes, D. Frishman, S. Stocker, A.N. Lupas, W. Baumeister, The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*, Nature 407 (2000) 508–513.

[32] S. Shigenobu, H. Watanabe, M. Hattori, Y. Sakaki, H. Ishikawa, Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS, Nature 407 (2000) 81–86.

[33] M. Shirai, H. Hirakawa, M. Kimoto, M. Tabuchi, F. Kishi, K. Ouchi, T. Shiba, K. Ishii, M. Hattori, S. Kuhara, T. Nakazawa, Comparison of whole genome sequences of *Chlamydia pneumoniae* J138 from Japan and CWL029 from USA, Nucleic Acids Res. 28 (2000) 2311–2314.

[34] A.J. Simpson, F.c. Reinach, P. Arruda, F.A. Abreu, M. Acencio, R. Alvarenga, L.M.C. Alves, J.E. Araya, G.S. Baia, C.S. Baptista, M.H. Barros, E.D. Bonaccorsi, S. Bordin, J.M. Bové, M.R.S. Briones, M.R.P. Bueno, A.A. Camargo, L.E.A. Camargo, D.M. Carraro, H. Carrer, N.B. Colauto, C. Colombo, F.F. Costa, M.C.R. Costa, C.M. Costa-neto et al., The genome sequence of the plant pathogen *Xylella fastidiosa*, Nature 406 (2000) 151–157.

[35] C.K. Stover, X.Q. Pham, A.L. Erwin, S.D. Mizoguchi, P. Warrener, M.J. Hickey, F.S. Brinkman, W.O. Hufnagle,

D.J. Kowalik, M. Lagrou, R.L. Garber, L. Goltry, E. Tolentino, S. Westbrock-Wadman, Y. Yuan, L.L. Brody, S.N. Coulter, K.R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G.K. Wong, Z. Wu et al., Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen, Nature 406 (2000) 959–964.

[36] H. Takami, K. Nakasone, Y. Takaki, G. Maeno, R. Sasaki, N. Masui, F. Fuji, C. Hirama, Y. Nakamura, N. Ogasawara, S. Kuhara, K. Horikoshi, Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*, Nucleic Acids Res. 28 (2000) 4317–4331.

[37] H. Tettelin, N.J. Saunders, J. Heidelberg, A.C. Jeffries, K.E. Nelson, J.A. Eisen, K.A. Ketchum, D.W. Hood, J.F. Peden, R.J. Dodson, W.C. Nelson, M.L. Gwinn, R. DeBoy, J.D. Peterson, E.K. Hickey, D.H. Haft, S.L. Salzberg, O. White, R.D. Fleischmann, B.A. Dougherty, T. Mason, A. Ciecko, D.S. Parksey, E. Blair, H. Cittone et al., Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58, Science 287 (2000) 1809–1815.

[38] S.T. Cole, K. Eiglmeier, J. Parkhill, K.D. James, N.R. Thomson, P.R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R.M. Davies, K. Devlin, S. Duthoy, T. Feltwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix et al., Massive gene decay in the leprosy bacillus, Nature 409 (2001) 1007–1011.

[39] B.J. May, Q. Zhang, L.L. Li, M.L. Paustian, T.S. Whittam, V. Kapur, Complete genomic sequence of *Pasteurella multocida*, Pm70, Proc. Natl. Acad. Sci. USA 98 (2001) 3460–3465.

[40] N.T. Perna, G. Plunkett III, V. Burland, B. Mau, J.D. Glasner, D.J. Rose, G.F. Mayhew, P.S. Evans, J. Gregor, H.A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E.J. Grotbeck, N.W. Davis, A. Lim, E.T. Dimalanta, K.D. Potamousis, J. Apodaca, T.S. Anantharaman, J. Lin, G. Yen, D.C. Schwartz, R.A. Welch, F.R. Blattner, Genome sequence of enterohaemorrhagic *Escherichia coli* O157H7, Nature 409 (2001) 529–533.

[41] R.L. Tatusov, E.V. Koonin, D.J. Lipman, A genomic perspective on protein families, Science 278 (1997) 631–637.

[42] H. Bono, H. Ogata, S. Goto, M. Kanehisa, Reconstruction of amino acid biosynthesis pathways from the complete genome sequence, Genome Res. 8 (1998) 203–210.

[43] T. Dandekar, S. Schuster, B. Snel, M. Huynen, P. Bork, Pathway alignment: application to the comparative analysis of glycolytic enzymes, Biochem. J. 343 (1999) 115–124.

[44] M.A. Huynen, T. Dandekar, P. Bork, Variation and evolution of the citric-acid cycle: a genomic perspective, Trends Microbiol. 7 (1999) 281–291.

[45] C.A. Harrington, C. Rosenow, J. Retief, Monitoring gene expression using DNA microarrays, Curr. Opin. Microbiol. 3 (2000) 285–291.

[46] M.P. Washburn, J.R. Yates, Analysis of the microbial proteome, Curr. Opin. Microbiol. 3 (2000) 292–297.

[47] T. Fenchel, B.J. Finlay, Ecology and Evolution in Anoxic Worlds, Oxford University Press, Oxford, 1995.

[48] D.R. Lovley, J.D. Coates, Novel forms of anaerobic respiration of environmental relevance, Curr. Opin. Microbiol. 3 (2000) 252–256.

[49] J.M. Shively, R.S. English, S.H. Baker, G.C. Cannon, Carbon cycling: the prokaryotic contribution, Curr. Opin. Microbiol. 4 (2001) 301–306.

[50] R.W. Ye, S.M. Thomas, Microbial nitrogen cycles: physiology, genomics and applications, Curr. Opin. Microbiol. 4 (2001) 307–312.

[51] J.F. Banfield, C.R. Marshall, Earth science and evolution genomics and the geosciences, Science 287 (2000) 605–606.

[52] M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30.

[53] R.L. Tatusov, D.A. Natale, I.V. Garkavtsev, T.A. Tatusova, U.T. Shankavaram, B.S. Rao, B. Kiryutin, M.Y. Galperin, N.D. Fedorova, E.V. Koonin, The COG database: new developments in phylogenetic classification of proteins from complete genomes, Nucleic Acids Res. 29 (2001) 22–28.

[54] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.

[55] B.L. Maidak, J.R. Cole, T.G. Lilburn, C.T. Parker Jr., P.R. Saxman, R.J. Farris, G.M. Garrity, G.J. Olsen, T.M. Schmidt, J.M. Tiedje, The RDP-II (Ribosomal Database Project), Nucleic Acids Res. 29 (2001) 173–174.

[56] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.

[57] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol. Biol. Evol. 17 (2000) 540–552.

[58] K. Adachi, M. Hasegawa, MOLPHY version 2.3: Programs for molecular phylogenetics based on maximum likelihood, Comput. Sci. Monogr. 28 (1996) 1–150.

[59] D.L. Swofford, PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4, Sinauer Associates, Sunderland, MA, 1998.

[60] C.R. Woese, O. Kandler, M.L. Wheelis, Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya, Proc. Natl. Acad. Sci. USA 87 (1990) 4576–4579.

[61] D. Yang, Y. Oyaizu, H. Oyaizu, G.J. Olsen, C.R. Woese, Mitochondrial origins, Proc. Natl. Acad. Sci. USA 82 (1985) 4443–4447.

[62] M.W. Gray, The endosymbiont hypothesis revisited, Int. Rev. Cytol. 141 (1992) 233–357.

[63] M.W. Gray, Evolution of organellar genomes, Curr. Opin. Genet. Dev. 9 (1999) 678–687.

[64] C.R. Woese, Bacterial evolution, Microbiol. Rev. 51 (1987) 221–271.

[65] D.M. Hillis, M.T. Dixon, Ribosomal DNA: molecular evo-

lution and phylogenetic inference, Q. Rev. Biol. 66 (1991) 411–453.
[66] G.J. Olsen, C.R. Woese, Ribosomal RNA: a key to phylogeny, FASEB J. 7 (1993) 113–123.
[67] Y. Van de Peer, P. De Rijk, J. Wuyts, T. Winkelmans, R. De Wachter, The European small subunit ribosomal RNA database, Nucleic Acids Res. 28 (2000) 175–176.
[68] J.P. Gogarten, H. Kibak, P. Dittrich, L. Taiz, E.J. Bowman, B.J. Bowman, M.F. Manolson, R.J. Poole, T. Date, T. Oshima, K.J.K. Denda, M. Yoshida, Evolution of the vacuolar H^+-ATPase: implications for the origin of eukaryotes, Proc. Natl. Acad. Sci. USA 86 (1989) 6661–6665.
[69] N. Iwabe, K. Kuma, M. Hasegawa, S. Osawa, T. Miyata, Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes, Proc. Natl. Acad. Sci. USA 86 (1989) 9355–9359.
[70] J.R. Brown, W.F. Doolittle, Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications, Proc. Natl. Acad. Sci. USA 92 (1995) 2441–2445.
[71] P. Forterre, H. Philippe, Where is the root of the universal tree of life?, Bioessays 21 (1999) 871–879.
[72] P. Lopez, P. Forterre, H. Philippe, The root of the tree of life in the light of the covarion model, J. Mol. Evol. 49 (1999) 496–508.
[73] H. Philippe, P. Forterre, The rooting of the universal tree of life is not reliable, J. Mol. Evol. 49 (1999) 509–523.
[74] A. Balows, H.G. Trüper, M. Dworkin, W. Harder, K.H. Schleifer, The Prokaryotes, 2nd edn., Springer, New York, 1992.
[75] G.J. Olsen, C.R. Woese, R. Overbeek, The winds of (evolutionary) change: breathing new life into microbiology, J. Bacteriol. 176 (1994) 1–6.
[76] S. Burggraf, G.J. Olsen, K.O. Stetter, C.R. Woese, A phylogenetic analysis of *Aquifex pyrophilus*, Syst. Appl. Microbiol. 15 (1992) 352–356.
[77] M. Bocchetta, E. Ceccarelli, R. Creti, A.M. Sanangelantoni, O. Tiboni, P. Cammarano, Arrangement and nucleotide sequence of the gene (*fus*) encoding elongation factor G (EF-G) from the hyperthermophilic bacterium *Aquifex pyrophilus*: phylogenetic depth of hyperthermophilic bacteria inferred from analysis of the EF-G/*fus* sequences, J. Mol. Evol. 41 (1995) 803–812.
[78] M. Saraste, Structural features of cytochrome oxidase, Q. Rev. Biophys. 23 (1990) 331–366.
[79] M. Saraste, L. Holm, L. Lemieux, M. Lübben, J. van der Oost, The happy family of cytochrome oxidases, Biochem. Soc. Trans. 19 (1991) 608–612.
[80] J.A. Garcia-Horsman, B. Barquera, J. Rumbley, J. Ma, R.B. Gennis, The superfamily of heme-copper respiratory oxidases, J. Bacteriol. 176 (1994) 5587–5600.
[81] M. Saraste, J. Castresana, D.G. Higgins, M. Lübben, M. Wilmanns, Evolution of cytochrome oxidase, in: H. Baltscheffsky (Ed.), Origin and Evolution of Biological Energy Conversion, VCH, New York, 1996, pp. 255–289.
[82] J. Castresana, M. Lübben, M. Saraste, D.G. Higgins, Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen, EMBO J. 13 (1994) 2516–2525.
[83] J. Castresana, M. Saraste, Evolution of energetic metabolism: the respiration-early hypothesis, Trends Biochem. Sci. 20 (1995) 443–448.
[84] M. Lübben, B. Kolmerer, M. Saraste, An archaebacterial terminal oxidase combines core structures of two mitochondrial respiratory complexes, EMBO J. 11 (1992) 805–812.
[85] M. Lübben, S. Arnaud, J. Castresana, A. Warne, S.P. Albracht, M. Saraste, A second terminal oxidase in *Sulfolobus acidocaldarius*, Eur. J. Biochem. 224 (1994) 151–159.
[86] O. Preisig, D. Anthamatten, H. Hennecke, Genes for a microaerobically induced oxidase complex in *Bradyrhizobium japonicum* are essential for a nitrogen-fixing endosymbiosis, Proc. Natl. Acad. Sci. USA 90 (1993) 3309–3313.
[87] R.K. Poole, G.M. Cook, Redundancy of aerobic respiratory chains in bacteria? Routes, reasons and regulation, Adv. Microb. Physiol. 43 (2000) 165–224.
[88] S. Jünemann, Cytochrome *bd* terminal oxidase, Biochim. Biophys. Acta 1321 (1997) 107–127.
[89] J.P. Osborne, R.B. Gennis, Sequence analysis of cytochrome *bd* oxidase suggests a revised topology for subunit I, Biochim. Biophys. Acta 1410 (1999) 32–50.
[90] J. Sakamoto, E. Koga, T. Mizuta, C. Sato, S. Noguchi, N. Sone, Gene structure and quinol oxidase activity of a cytochrome *bd*-type oxidase from *Bacillus stearothermophilus*, Biochim. Biophys. Acta 1411 (1999) 147–158.
[91] W.G. Zumft, The denitrifying prokaryotes, in: A. Balows, H.G. Trüper, M. Dworkin, W. Harder, K.H. Schleifer (Eds.), The Prokaryotes, 2nd edn., Springer, New York, 1992, pp. 554–582.
[92] W.G. Zumft, Cell biology and molecular basis of denitrification, Microbiol. Mol. Biol. Rev. 61 (1997) 533–616.
[93] D.J. Richardson, Bacterial respiration: a flexible process for a changing environment, Microbiology 146 (2000) 551–571.
[94] D.J. Richardson, B.C. Berks, D.A. Russell, S. Spiro, C.J. Taylor, Functional, biochemical and genetic diversity of prokaryotic nitrate reductases, Cell. Mol. Life Sci. 58 (2001) 165–178.
[95] J. Hendriks, A. Oubrie, J. Castresana, A. Urbani, S. Gemeinhardt, M. Saraste, Nitric oxide reductases in bacteria, Biochim. Biophys. Acta 1459 (2000) 266–273.
[96] N. Speich, C. Dahl, P. Heisig, A. Klein, F. Lottspeich, K.O. Stetter, H.G. Trüper, Adenylylsulphate reductase from the sulphate-reducing archaeon *Archaeoglobus fulgidus*: cloning and characterization of the genes and comparison of the enzyme with other iron–sulphur flavoproteins, Microbiology 140 (1994) 1273–1284.
[97] W.M. Hipp, A.S. Pott, N. Thum-Schmitz, I. Faath, C. Dahl, H.G. Trüper, Towards the phylogeny of APS reductases and sirohaem sulfite reductases in sulfate-reducing and sulfur-oxidizing prokaryotes, Microbiology 143 (1997) 2891–2902.
[98] T. Krafft, M. Bokranz, O. Klimmek, I. Schroder, F. Fahrenholz, E. Kojro, A. Kroger, Cloning and nucleotide se-

quence of the *psrA* gene of *Wolinella succinogenes* polysulphide reductase, Eur. J. Biochem. 206 (1992) 503–510.
[99] J. Castresana, D. Moreira, Respiratory chains in the last common ancestor of living organisms, J. Mol. Evol. 49 (1999) 453–460.
[100] C. Dahl, N.M. Kredich, R. Deutzmann, H.G. Trüper, Dissimilatory sulphite reductase from *Archaeoglobus fulgidus*: physico-chemical properties of the enzyme and cloning, sequencing and analysis of the reductase genes, J. Gen. Microbiol. 139 (1993) 1817–1828.
[101] M. Wagner, A.J. Roger, J.L. Flax, G.A. Brusseau, D.A. Stahl, Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration, J. Bacteriol. 180 (1998) 2975–2982.
[102] M. Klein, M. Friedrich, A.J. Roger, P. Hugenholtz, S. Fishbain, H. Abicht, L.L. Blackall, D.A. Stahl, M. Wagner, Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes, J. Bacteriol. 183 (2001) 6028–6035.
[103] J.N. Reeve, J. Nolling, R.M. Morgan, D.R. Smith, Methanogenesis: genes, genomes, and who's on first?, J. Bacteriol. 179 (1997) 5975–5986.
[104] R.K. Thauer, Biochemistry of methanogenesis: a tribute to Marjory Stephenson, Microbiology 144 (1998) 2377–2406.
[105] W. Nitschke, A.W. Rutherford, Photosynthetic reaction centres: variations on a common structural theme?, Trends Biochem. Sci. 16 (1991) 241–245.
[106] R.E. Blankenship, H. Hartman, The origin and evolution of oxygenic photosynthesis, Trends Biochem. Sci. 23 (1998) 94–97.
[107] P. Schönheit, T. Schäfer, Metabolism of hyperthermophiles, World J. Microbiol. Biotechnol. 11 (1995) 26–57.
[108] M.T. Madigan, J.M. Martinko, J. Parker, Brock Biology of Microorganisms, 8th edn., Prentice Hall, Upper Saddle River, NJ, 1997.
[109] A.H. Stouthamer, Metabolic pathways in *Paracoccus denitrificans* and closely related bacteria in relation to the phylogeny of prokaryotes, Antonie Van Leeuwenhoek 61 (1992) 1–33.
[110] N. Glansdorff, About the last common ancestor, the universal life-tree and lateral gene transfer: a reappraisal, Mol. Microbiol. 38 (2000) 177–185.
[111] S.G. Andersson, C.G. Kurland, Reductive evolution of resident genomes, Trends Microbiol. 6 (1998) 263–268.
[112] H. Ochman, N.A. Moran, Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis, Science 292 (2001) 1096–1098.
[113] M. Riley, Functions of the gene products of *Escherichia coli*, Microbiol. Rev. 57 (1993) 862–952.
[114] M. Riley, Systems for categorizing functions of gene products, Curr. Opin. Struct. Biol. 8 (1998) 388–392.
[115] J.R. Brown, W.F. Doolittle, Archaea and the prokaryote-to-eukaryote transition, Microbiol. Mol. Biol. Rev. 61 (1997) 456–502.
[116] M.W. Smith, D.F. Feng, R.F. Doolittle, Evolution by acquisition: the case for horizontal gene transfers, Trends Biochem. Sci. 17 (1992) 489–493.
[117] M. Syvanen, Horizontal gene transfer: evidence and possible consequences, Annu. Rev. Genet. 28 (1994) 237–261.
[118] W.F. Doolittle, Lateral genomics, Trends Cell Biol. 9 (1999) M5–M8.
[119] P. Forterre, C. Bouthier De La Tour, H. Philippe, M. Duguet, Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from archaea to bacteria, Trends Genet. 16 (2000) 152–154.
[120] H. Ochman, J.G. Lawrence, E.A. Groisman, Lateral gene transfer and the nature of bacterial innovation, Nature 405 (2000) 299–304.
[121] N.C. Kyrpides, G.J. Olsen, Archaeal and bacterial hyperthermophiles: horizontal gene exchange or common ancestry?, Trends Genet. 15 (1999) 298–299.
[122] C.G. Kurland, Something for everyone: Horizontal gene transfer in evolution, EMBO Rep. 1 (2000) 92–95.
[123] S.L. Salzberg, O. White, J. Peterson, J.A. Eisen, Microbial genes in the human genome: lateral transfer or gene loss?, Science 292 (2001) 1903–1906.
[124] M.J. Stanhope, A. Lupas, M.J. Italia, K.K. Koretke, C. Volker, J.R. Brown, Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates, Nature 411 (2001) 940–944.
[125] K.H. Wolfe, D.C. Shields, Molecular evidence for an ancient duplication of the entire yeast genome, Nature 387 (1997) 708–713.
[126] M. Lynch, J.S. Conery, The evolutionary fate and consequences of duplicate genes, Science 290 (2000) 1151–1155.
[127] K.H. Wolfe, Yesterday's polyploids and the mystery of diploidization, Nat. Rev. Genet. 2 (2001) 333–341.
[128] P. Forterre, N. Benachenhou-Lahfa, F. Confalonieri, M. Duguet, C. Elie, B. Labedan, The nature of the last universal ancestor and the root of the tree of life, still open questions, BioSystems 28 (1993) 15–32.
[129] C.R. Woese, The universal ancestor, Proc. Natl. Acad. Sci. USA 95 (1998) 6854–6859.
[130] N. Kyrpides, R. Overbeek, C. Ouzounis, Universal protein families and the functional content of the last universal common ancestor, J. Mol. Evol. 49 (1999) 413–423.
[131] D. Penny, A. Poole, The nature of the last universal common ancestor, Curr. Opin. Genet. Dev. 9 (1999) 672–677.
[132] W.F. Doolittle, The nature of the universal ancestor and the evolution of the proteome, Curr. Opin. Struct. Biol. 10 (2000) 355–358.
[133] R. Guigo, I. Muchnik, T.F. Smith, Reconstruction of ancient molecular phylogeny, Mol. Phylogenet. Evol. 6 (1996) 189–213.
[134] J. Castresana, M. Lübben, M. Saraste, New archaebacterial genes coding for redox proteins: implications for the evolution of aerobic metabolism, J. Mol. Biol. 250 (1995) 202–210.
[135] K.M. Towe, Earth's early atmosphere: constraints and opportunities for early evolution, in: S. Bengtson (Ed.), Early

Life on Earth, Columbia University Press, New York, 1994, pp. 36–47.
[136] K.M. Towe, Environmental oxygen conditions during the origin and early evolution of life, Adv. Space Res. 18 (1996) 7–15.
[137] G. Schäfer, W.G. Purschke, M. Gleissner, C.L. Schmidt, Respiratory chains of archaea and extremophiles, Biochim. Biophys. Acta 1275 (1996) 16–20.
[138] M. Gleissner, M.G. Elferink, A.J. Driessen, W.N. Konings, S. Anemüller, G. Schäfer, Generation of proton-motive force by an archaeal terminal quinol oxidase from *Sulfolobus acidocaldarius*, Eur. J. Biochem. 224 (1994) 983–990.
[139] M.G. Elferink, J.G. De Wit, A.J. Driessen, W.N. Konings, Energy-transducing properties of primary proton pumps reconstituted into archaeal bipolar lipid vesicles, Eur. J. Biochem. 214 (1993) 917–925.
[140] M. Saraste, J. Castresana, Cytochrome oxidase evolved by tinkering with denitrification enzymes, FEBS Lett. 341 (1994) 1–4.
[141] J. van der Oost, A.P. de Boer, J.W. de Gier, W.G. Zumft, A.H. Stouthamer, R.J. van Spanning, The heme-copper oxidase family consists of three distinct types of terminal oxidases and is related to nitric oxide reductase, FEMS Microbiol. Lett. 121 (1994) 1–9.
[142] A. Giuffre, G. Stubauer, P. Sarti, M. Brunori, W.G. Zumft, G. Buse, T. Soulimane, The heme-copper oxidases of *Thermus thermophilus* catalyze the reduction of nitric oxide: evolutionary implications, Proc. Natl. Acad. Sci. USA 96 (1999) 14718–14723.
[143] M.W. Gray, G. Burger, B.F. Lang, Mitochondrial evolution, Science 283 (1999) 1476–1481.
[144] O. Karlberg, B. Canbäck, C.G. Kurland, S.G. Andersson, The dual origin of the yeast mitochondrial proteome, Yeast 17 (2000) 170–187.
[145] C.G. Kurland, S.G. Andersson, Origin and evolution of the mitochondrial proteome, Microbiol. Mol. Biol. Rev. 64 (2000) 786–820.