The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial megaclassification

Department of Zoology,
University of Oxford,
South Parks Road, Oxford
OX1 3PS, UK

T. Cavalier-Smith

Tel: +44 1865 281065. Fax: +44 1865 281310. e-mail: tom.cavalier-smith@zoo.ox.ac.uk

Prokaryotes constitute a single kingdom, Bacteria, here divided into two new subkingdoms: Negibacteria, with a cell envelope of two distinct genetic membranes, and Unibacteria, comprising the new phyla Archaeobacteria and Posibacteria, with only one. Other new bacterial taxa are established in a revised higher-level classification that recognizes only eight phyla and 29 classes. Morphological, palaeontological and molecular data are integrated into a unified picture of large-scale bacterial cell evolution despite occasional lateral gene transfers. Archaeobacteria and eukaryotes comprise the clade neomura, with many common characters, notably obligately co-translational secretion of *N*-linked glycoproteins, signal recognition particle with 7S RNA and translation-arrest domain, protein-spliced tRNA introns, eight-subunit chaperonin, prefoldin, core histones, small nucleolar ribonucleoproteins (snRNPs), exosomes and similar replication, repair, transcription and translation machinery. Eubacteria (posibacteria and negibacteria) are paraphyletic, neomura having arisen from Posibacteria within the new subphylum Actinobacteria (possibly from the new class Arabobacteria, from which eukaryotic cholesterol biosynthesis probably came). Replacement of eubacterial peptidoglycan by glycoproteins and adaptation to thermophily are the keys to neomuran origins. All 19 common neomuran character suites probably arose essentially simultaneously during the radical modification of an actinobacterium. At least 11 were arguably adaptations to thermophily. Most unique archaeobacterial characters (prenyl ether lipids; flagellar shaft of glycoprotein, not flagellin; DNA-binding protein 10b; specially modified tRNA; absence of Hsp90) were subsequent secondary adaptations to hyperthermophily and/or hyperacidity. The insertional origin of protein-spliced tRNA introns and an insertion in proton-pumping ATPase also support the origin of neomura from eubacteria. Molecular co-evolution between histones and DNA-handling proteins, and in novel protein initiation and secretion machineries, caused quantum evolutionary shifts in their properties in stem neomura. Proteasomes probably arose in the immediate common ancestor of neomura and Actinobacteria. Major gene losses (e.g. peptidoglycan synthesis, *hsp90*, *secA*) and genomic reduction were central to the origin of archaeobacteria. Ancestral archaeobacteria were probably heterotrophic, anaerobic, sulphur-dependent hyperthermoacidophiles; methanogenesis and halophily are secondarily derived. Multiple lateral gene transfers from eubacteria helped secondary archaeobacterial adaptations to mesophily and genome re-expansion. The origin from a drastically altered actinobacterium of neomura, and the immediately subsequent simultaneous origins of archaeobacteria and eukaryotes, are the most extreme and important cases of

This paper is an elaboration of part of an invited presentation to the XIIIth meeting of the International Society for Evolutionary Protistology in České Budějovice, Czech Republic, 31 July–4 August 2000.

Abbreviations: ER, endoplasmic reticulum; GlcNac, *N*-acetylglucosamine; RuBisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; snRNP, small nucleolar ribonucleoprotein; TCA, tricarboxylic acid.

quantum evolution since cells began. All three strikingly exemplify De Beer's principle of mosaic evolution: the fact that, during major evolutionary transformations, some organismal characters are highly innovative and change remarkably swiftly, whereas others are largely static, remaining conservatively ancestral in nature. This phenotypic mosaicism creates character distributions among taxa that are puzzling to those mistakenly expecting uniform evolutionary rates among characters and lineages. The mixture of novel (neomuran or archaeobacterial) and ancestral eubacteria-like characters in archaeobacteria primarily reflects such vertical mosaic evolution, not chimaeric evolution by lateral gene transfer. No symbiogenesis occurred. Quantum evolution of the basic neomuran characters, and between sister paralogues in gene duplication trees, makes many sequence trees exaggerate greatly the apparent age of archaeobacteria. Fossil evidence is compelling for the extreme antiquity of eubacteria [over 3500 million years (My)] but, like their eukaryote sisters, archaeobacteria probably arose only 850 My ago. Negibacteria are the most ancient, radiating rapidly into six phyla. Evidence from molecular sequences, ultrastructure, evolution of photosynthesis, envelope structure and chemistry and motility mechanisms fits the view that the cenancestral cell was a photosynthetic negibacterium, specifically an anaerobic green non-sulphur bacterium, and that the universal tree is rooted at the divergence between sulphur and non-sulphur green bacteria. The negibacterial outer membrane was lost once only in the history of life, when Posibacteria arose about 2800 My ago after their ancestors diverged from Cyanobacteria.

Keywords: Unibacteria, Actinobacteria, thermophily and molecular co-evolution of DNA-handling enzymes, origin of *N*-linked glycoprotein secretion, microbial fossils and evolution

Introduction and overview

Recent genome sequencing has fostered a simplistic view of organisms as essentially aggregates of genes. However, organisms are not simply a sum of their genes nor, as some biochemists were once wont to say, mere bags of enzymes. Genes and enzymes are both fundamental, but play their vital roles as parts of highly organized growing and dividing cells. Their life depends on a mutualistic symbiosis of genes, catalysts, membranes and cell skeleton (Cavalier-Smith, 1987a, 1991a, b, 2001). Co-adaptation between co-operating not selfish molecules is the key to understanding living organisms. The degree to which different cellular macromolecules are co-adapted varies greatly; for many metabolic enzymes, direct co-adaptation in structure is low, integration being mediated through non-informational intermediary metabolites, but for many informational and structural molecules it is high. Genetic information is made manifest through physical structure. DNA is physically inert – genes do not make organisms; they grow by physico-chemical interactions between effector macromolecules whose structure and physico-chemical properties are genetically determined. Membranes of lipids with embedded proteins are centrally important: chromosomes, ribosomes and the cytoskeleton physically attach to them;

the cell's structural integrity and its character as a growing and reproducing organism depend on these direct physical interconnections. The ability of membranes to sequester food, grow and divide underlies cell growth and reproduction. Like chromosomes, but unlike ribosomes and the skeleton, membranes show direct genetic continuity: all are descended by growth and division from those bounding the first cell (Cavalier-Smith, 1991a, b). Membranes have a hereditary role as well as structural and physiological roles (Cavalier-Smith, 2000a, 2001). The unity of life stems from the common origin and fundamental similarity of these processes in all organisms. Organismal structural diversity, on the other hand, arises through variations in membrane topology and physico-chemical properties as well as in the shapes formed by the cell skeleton, for both of which the genically specified catalysts create the building blocks. This means that we cannot understand the evolution of life without elucidating the evolution of cell organization and reproduction as well as that of the individual molecules that mediate them.

The most profound difference within the living world lies between bacteria and eukaryotes (Stanier & Van Niel, 1962; Stanier, 1970; Cavalier-Smith, 1987b, 1991a, b, 1998). 'Bacteria' in this paper is used in the

proper traditional sense to embrace all prokaryotes (Cavalier-Smith, 1992b, 1998; Mayr, 1998), never as a fashionable but highly confusing synonym for eubacteria only (Woese *et al.*, 1990). Bacteria and eukaryotes differ fundamentally in the topological relationships between membranes, genomes and ribosomes and in their skeletons. In all bacteria, chromosomal DNA and ribosomes making membrane proteins are attached directly to the cytoplasmic membrane, which grows by the direct insertion of proteins and lipids. In eukaryotes, the chromosomes and ribosomes making membrane proteins are attached instead to the endoplasmic reticulum (ER)/nuclear envelope, which is topologically within, and unconnected to, the plasma membrane, which grows by fusion of vesicles budded from endomembranes; the ER grows, like the bacterial cytoplasmic membrane, by the direct insertion of individual lipid molecules synthesized by proteins embedded within the same membrane. All eukaryotes have a complex endoskeleton (the cytoskeleton) of microtubules and actin filaments that use attached molecular motors to mediate chromosome segregation and cell division, respectively. By contrast, bacteria have an exoskeleton (cell wall) important for DNA segregation and cell division. There has been much discussion of how these and other profound differences between bacteria and eukaryotes have arisen (Margulis, 1970; Cavalier-Smith, 1975, 1980, 1981, 1987b, 1990, 1991a, b, c, 1992c, 1993, 2000b; de Duve, 1996; Faguy & Doolittle, 1998), updated in a following paper (Cavalier-Smith, 2002). The primary purpose of this paper is to discuss the origins of the less profound, but highly important differences between the three major types of bacteria: the structurally simple archaeobacteria (Woese & Fox, 1977) and posibacteria (Cavalier-Smith, 1987b) and the topologically more complex negibacteria (Cavalier-Smith, 1987b).

Archaeobacteria and posibacteria are bounded by a single membrane only and are thus referred to collectively as unibacteria (Cavalier-Smith, 1998). Negibacteria, in sharp contrast, are bounded by two topologically distinct membranes; the cytoplasmic membrane, into which lipids and proteins are inserted directly, and the relatively porous outer membrane that grows more indirectly by their subsequent transfer across specific adhesion sites between the two. The biogenesis of the negibacterial envelope is more complex and requires extra chaperones. As the cytoplasmic membrane of posibacteria and negibacteria is composed of acyl ester lipids, like eukaryotic membranes, they are grouped together as eubacteria, so as to contrast them with archaeobacteria, which are unique in the living world in having prenyl ether lipids instead.

Two fundamentally different views have been proposed of the significance of this and other striking differences between archaeobacteria and eubacteria. One influential school of thought regards them as ancient differences that reflect an early divergence soon after the origin of life, before many cell characters had

become stabilized (Woese & Fox, 1977; Woese, 1998, 2000; Graham *et al.*, 2000). The second view is that archaeobacteria are not an ancient group at all (Hori *et al.*, 1982) but arose secondarily from eubacteria relatively recently as an adaptation to hyperthermophily (Cavalier-Smith, 1987a, b, 1991a, b, 1998; Forterre, 1996); although not all archaeobacteria are thermophiles, it is argued that their last common ancestor was a hyperthermophile and that it arose from a eubacterial ancestor by lipid replacement and other adaptations. Here, I review recent evidence and arguments that, in my view, support compellingly the secondarily derived nature of archaeobacteria. It is now well established that archaeobacteria are either ancestral to (Van Valen & Maiorana, 1980) or, more likely (Cavalier-Smith, 1987b), sisters of eukaryotes, with which they share many important characters. When first proposing that archaeobacteria and eukaryotes were sister taxa, I called the clade that comprised them *neomura* (new walls), because I considered that their shared *N*-linked glycoproteins were derived compared with the ancestral peptidoglycans of eubacteria, arguing that the fossil record implied that *neomura* were less than half the age of eubacteria (Cavalier-Smith, 1987b). I also asserted that *neomura* evolved from posibacteria by the replacement of peptidoglycan by *N*-linked glycoproteins and tentatively suggested that *neomura* are more closely related to high-G + C Gram-positive bacteria (the subphylum Actinobacteria) than to low-G + C Gram-positives (here collectively grouped with mycoplasmas and their heliobacterial and thermotogalean allies as a new subphylum, Endobacteria).

This paper reviews recent evidence that very strongly supports such an actinobacterial origin for the *neomura* and develops the secondary hyperthermophily hypothesis of the origin of archaeobacteria (Cavalier-Smith, 1987a, b) in more detail. A critical re-evaluation of the fossil record in the present paper indicates that eukaryotes are much younger than often thought (Cavalier-Smith, 1990) – probably only about 850 million years (My) old. The bacterial fossil record clearly indicates that eubacteria are far more ancient, at least 3500 My old. Dating archaeobacterial origins is more problematic, but I shall argue that, like eukaryotes, they are probably at least four times younger than eubacteria. The present paper also severely criticizes arguments and assumptions that have been used to suggest that archaeobacteria and/or eukaryotes may be more ancient than or as old as eubacteria. The somewhat revised classification of the kingdom Bacteria adopted here is summarized in Table 1; my reasons for treating all prokaryotes as a single kingdom Bacteria, and why eubacteria are not a clade and are preferably not treated as a taxon, were explained previously (Cavalier-Smith, 1998).

My arguments that *neomuran* and archaeobacterial characteristics are all relatively recently derived characters in no way trivializes the importance of the numerous differences between archaeobacteria and

Table 1. Revised classification of kingdom Bacteria and its eight phyla (= divisions)

Revised from Cavalier-Smith (1992a, 1998); the latter includes a formal description of the kingdom Bacteria: I here validate it under the Bacteriological Code by designating Enterobacteriales as the type order. Eubacteria is a useful grade name, but is not treated as a taxon here.

Taxon	Etymology	Description	Type
Subkingdom 1. NEGIBACTERIA* (Cavalier-Smith, 1987b) subregnum nov.	Contraction from <i>L. negativus</i> negative, since most stain Gram-negative	Cell bounded by two concentric lipid bilayers, the cytoplasmic membrane and an outer membrane bearing porins; ancestrally with peptidoglycan and lipoprotein between the membranes; SRP lacks helices 1–4 and 19p; protein secretion predominantly post-translational	Order Enterobacteriales
Infrakingdom 1. Eobacteria (Cavalier-Smith, 1992a) infraregnum nov.	Gr. <i>eos</i> dawn, because the absence of lipopolysaccharide suggests they may be the earliest negibacteria	No lipopolysaccharide or sphingolipids; peptidoglycan with ornithine, not diaminopimelic acid; usually thermophilic; flagella absent; gas vesicles absent	Order Chloroflexales
Division 1. Eobacteria (Cavalier-Smith, 1992a) divisio nov.	As for infrakingdom above	As for infrakingdom above	Order Chloroflexales
Class 1. Chlorobacteria (Cavalier-Smith, 1992a) classis nov.	Gr. <i>khloros</i> yellow green, from the colour of the photosynthetic species	Filamentous green bacteria, with bacteriochlorophyll <i>a</i> and usually chlorosomes, gliding green non-sulphur photosynthetic bacteria, with phaeophytin quinone type-2 reaction centres, with or without chlorosomes (<i>Chloroflexus</i> , <i>Heliothrix</i> , <i>Roseiflexus</i> , <i>Oscillochloris</i>), and their colourless relatives, e.g. <i>Thermomicrobium</i> , <i>Herpetosiphon</i> , <i>Thermoleophilum</i> , <i>Dehalococcoides</i> (a halorespirer)	Order Chloroflexales
Class 2. Hadobacteria (Cavalier-Smith, 1992a; emend. 1998) classis nov.	Gr. <i>hades</i> hell, because they can resist extremes of heat or radiation	Heterotrophic thermophiles or highly radiation-resistant bacteria with thick murein layer; with semi-crystalline <i>S</i> -layer, e.g. <i>Deinococcus</i> , <i>Thermus</i> , <i>Meiothermus</i> ; more closely related to each other on rRNA trees than to Chlorobacteria	Order Thermales
Infrakingdom 2. Glycobacteria* (Cavalier-Smith, 1998) infraregnum nov.	Gr. <i>glukus</i> sweet, because they have surface lipopolysaccharide	Outer membrane with lipopolysaccharide or lipooligosaccharide; peptidoglycan with diaminopimelic acid or ornithine; gas vesicles widespread	Order Enterobacteriales
Division 1. Cyanobacteria (Stanier 1974) nom. rev. (ex Stanier & Cohen-Bazire, 1977 as class)	Gr. <i>kuanos</i> blue-green, because of their common colour and the traditional name Cyanophyceae or blue-green algae	Oxygenic photosynthesis with chlorophyll <i>a</i> ; flagella absent; often glide; ancestrally with phycobilisomes, sometimes lost	Order Chroococcales
Subdivision 1. Gloeobacteria subdivisio nov.	From <i>Gloeobacter</i> , the only known genus	Without thylakoids	Order Gloeobacterales
Class 1. Gloeobacteria (Cavalier-Smith, 1998) classis nov.	As for subdivision above	As for subdivision above	Order Gloeobacterales
Order 1. Gloeobacterales ord. nov.	As for subdivision above	Having phycobilisomes but no thylakoids	Genus <i>Gloeobacter</i>
Subdivision 2. Phycobacteria (Cavalier-Smith, 1998) subdivisio nov.	Gr. <i>phukos</i> seaweed, because all the traditional blue-green algae and the prochlorophytes are included	With thylakoids; gliding motility by slime secretion; classical Cyanophyceae and prochlorophytes. The five traditional cyanobacterial orders, already valid under the Code of Botanical Nomenclature, are here also formally validated under the Bacteriological (= Prokaryotic) Code	Order Chroococcales
Class 1. Chroobacteria classis nov.	From the genus <i>Chroococcus</i>	Unicellular, palmelloid, colonial or with filaments lacking heterocysts	Order Chroococcales
Order 1. Chroococcales ord. nov.	As for class above	Unicellular and colonial (non-filamentous) cyanobacteria (with phycobilisomes and prochlorophytes with chlorophyll <i>b</i> instead	Genus <i>Chroococcus</i>

Table 1 (cont.)

Taxon	Etymology	Description	Type
Order 2. Pleurocapsales ord. nov.	From the genus <i>Pleurocapsa</i>	Colonial or filamentous, reproducing by intramural multiple fission to yield smaller unicellular dispersal stages	Genus <i>Pleurocapsa</i>
Order 3. Oscillatoriales ord. nov.	From the genus <i>Oscillatoria</i>	Unbranched linear filaments without heterocysts; cells typically shorter than broad	Genus <i>Oscillatoria</i>
Class 2. Hormogoneae (ex Thuret 1875) classis nov.	Gr. <i>hormos</i> cord; Gr. <i>gonos</i> offspring; N.L. <i>hormogonia</i> hormogonia, because they multiply by hormogonia	Filaments that multiply vegetatively by hormogonia; usually with heterocysts	Order Nostocales
Order 1. Nostocales ord. nov.	From the genus <i>Nostoc</i>	Unbranched filaments	Genus <i>Nostoc</i>
Order 2. Stigonematales ord. nov.	From the genus <i>Stigonema</i>	Branched filaments	Genus <i>Stigonema</i>
Division 2. Spirochaetae	From the genus <i>Spirochaeta</i>	With spiral flagella driven by a rotary motor having shafts within periplasmic space; cell corkscrews through semisolid media; outer membrane flexible, with lipooligosaccharide instead of lipopolysaccharide; organotrophs lacking photosynthesis	Order Spirochaetales
Class Spirochaetes	As for division above	As for division above (e.g. <i>Treponema</i> , <i>Borrelia</i> , <i>Leptospira</i> , <i>Leptonema</i>)	Order Spirochaetales
Division 3. Sphingobacteria (Cavalier-Smith, 1987a) divisio nov.	Gr. <i>sphiggo</i> strangle, because they have sphingolipids	Cytoplasmic membrane with sphingolipids; outer membrane with lipopolysaccharide; usually mesophilic; flagella absent	Order Cytophagales
Class 1. Flavobacteria (Cavalier-Smith, 1998) classis nov.	From the genus <i>Flavobacterium</i>	Aerobic heterotrophs e.g. Cytophagales (predatory), Flavobacteriaceae, Bacteroidaceae, <i>Fibrobacter</i> . Includes <i>Flavobacterium</i> and its relatives	Order Cytophagales
Class 2. Chlorobea classis nov.	From the genus <i>Chlorobium</i>	Anaerobic phototrophs with homomeric type 1 reaction centres and chlorosomes. Sole and type order Chlorobiales. Includes <i>Chlorobium</i> and all other green sulphur bacteria	Order Chlorobiales
Superdivision Exoflagellata superdivisio nov.	Gr. <i>exo</i> outside; L. <i>flagellum</i> whip, because they include all negibacteria with flagellar shafts outside the outer membrane	Negibacteria with flagellar shafts outside the outer membrane; no sphingolipid	Order Enterobacteriales
Division 1. Planctobacteria (Cavalier-Smith, 1987) divisio nov.	From Planctomycetales, the type and best-known free-living members	Negibacteria lacking peptidoglycan plus their closest relatives	Order Planctomycetales
Class 1. Planctomycea classis nov.	As for the division above	With protein walls but no peptidoglycan; free-living, often flagellate, aquatic heterotrophs with budding division (e.g. <i>Pirellula</i> , <i>Gemmata</i>)	Order Planctomycetales
Class 2. Verrucomicrobiae (Hedlund <i>et al.</i> , 1997)	From the genus <i>Verrucomicrobium</i>	Prosthecae, free-living bacteria with murein or intracellular parasites lacking it	Order Verrucomicrobiales
Class 3. Chlamydiae classis nov.	From the genus <i>Chlamydia</i>	Obligate intracellular energy parasites of eukaryotes that import all their ATP; flagella absent; peptidoglycan absent, walls of protein (<i>Chlamydia</i>)	Order Chlamydiales
Division 2. Proteobacteria (ex Stackebrandt <i>et al.</i> , 1986 as class) divisio nov.	L. and Gr. <i>Proteus</i> a sea god able to assume many shapes, referring to the great variety of bacteria included	Always with peptidoglycan and lipopolysaccharide; multifarious respiratory patterns: large insertion in RNA polymerase and Hsp70	Order Enterobacteriales
Subdivision 1. Rhodobacteria (Cavalier-Smith, 1987a) subdivisio nov.	Gr. <i>rhodon</i> rose, because all purple photosynthetic bacteria, many with names beginning <i>Rhodo</i> -, are included	Ancestrally phototrophs with heterodimeric type 2 reaction centres with bacteriochlorophyll <i>a</i> , <i>c</i> and <i>d</i> and carotenoids located in extensive tubular or flattened membrane invaginations; plus their organotrophic (heterotrophic or methylotrophic) descendants; usually with ubiquinone	Order Enterobacteriales
Class 1. Chromatibacteria (Cavalier-Smith, 1998) classis nov.	From the genus <i>Chromatium</i>	Purple sulphur bacteria and their colourless heterotrophic or methylotrophic descendants; often with both ubiquinones and menaquinones; i.e. β -proteobacteria, e.g. Neisseriaceae, <i>Spirillum</i> , <i>Rhodocyclus</i> , <i>Thiobacillus</i> , Alcaligenaceae, and γ -proteobacteria, e.g. Chromatiaceae, Pseudomonadaceae, Methylococcaceae, Vibrionaceae, Enterobacteriaceae (e.g. <i>Escherichia</i>)	Order Enterobacteriales

Table 1 (cont.)

Taxon	Etymology	Description	Type
Class 2. Alphabacteria (Cavalier-Smith, 1992a) classis nov.	Gr. <i>alpha</i> the letter a, so as to formalize their earlier informal designation as α -proteobacteria	Non-sulphur purple bacteria and their heterotrophic descendants; respirers with ubiquinones having 10 isoprenoid units, often facultative aerobes; e.g. Rhodospirillaceae, <i>Rhodobacter</i> , <i>Caulobacter</i> , Bartonellaceae, <i>Methylobacterium</i> , <i>Rhizobium</i> , <i>Hyphomicrobium</i> , Rickettsiales	Order Rickettsiales
Subdivision 2. Thiobacteria (Cavalier-Smith, 1998) subdivisio nov.	Gr. <i>thion</i> sulphur, because sulphate reduction might have been their ancestral phenotype	Non-photosynthetic relatives, possibly sisters, of Rhodobacteria; sulphate-reducing respirers and their organotrophic relatives; with menaquinones but not ubiquinone	Order Myxococcales
Class 1. Delabacteria (Cavalier-Smith, 1992a) classis nov.	Gr. <i>delta</i> the letter d, to formalize their customary designation as δ -proteobacteria	Anaerobic sulphate reducers or typically aerobic organotrophs or predators; e.g. <i>Desulfohalobacterium</i> , <i>Bedellovibrio</i> , Myxococcales (type order) myxobacteria; fruiting gliders	Order Myxococcales
Class 2. Epsilobacteria classis nov.	Gr. <i>epsilon</i> the letter epsilon, to formalize the customary designation of most members as ϵ -proteobacteria	Organotrophs, often parasitic e.g. <i>Helicobacter</i> , or hydrogen-oxidizing lithotrophs; hyperthermophiles with alkylether lipids or thermophiles lacking them (e.g. <i>Aquifex</i> , <i>Hydrogenobacter</i>)	Order Aquificales
Thiobacteria <i>incertae sedis</i> : <i>Thermodesulfobacterium</i>			
Subdivision 3. Geobacteria subdivisio nov.	Gr. <i>ge</i> earth, because many are abundant in soil	Non-photosynthetic anaerobic respirers and their fermenting descendants; often iron-reducers or oxidizers; rarely sulphate reducers	Order Geovibriales
Class 1. Ferrobacteria classis nov.	<i>L. ferrum</i> iron, as many reduce it	Geobacteria phylogenetically closer to <i>Geovibrio</i> than to <i>Acidobacterium</i> , <i>Flexistipes</i> / <i>Denitrovibrio</i> / <i>Deferribacter</i> / <i>Geovibrio</i> group; <i>Synergistes</i> ; <i>Nitrospira</i> / <i>Magnetobacterium</i> / <i>Leptospirillum</i> / <i>Thermodesulfovibrio</i> group	Order Geovibriales
Order 1. Geovibriales ord. nov.	From the genus <i>Geovibrio</i>	As for class above	Genus <i>Geovibrio</i>
Class 2. Acidobacteria classis nov.	From the genus <i>Acidobacterium</i>	Geobacteria phylogenetically closer to <i>Acidobacterium</i> than to <i>Geovibrio</i> ; e.g. <i>Acidobacterium</i> , <i>Holophaga</i> , <i>Geothrix</i>	Order Acidobacteriales
Order 1. Acidobacteriales ord. nov.	As for class above	As for class above	Genus <i>Acidobacterium</i>
Subkingdom 2. UNIBACTERIA* (Cavalier-Smith, 1998) subregnum nov.	<i>L. unus</i> one, referring to the always single bounding membrane, in contrast to the two membranes of Negibacteria	Cell bounded by only a single cytoplasmic membrane; commonly with an external proteinaceous paracrystalline S-layer; protein secretion predominantly co-translational	Order Bacillales
Division 1. Posibacteria* (Cavalier-Smith, 1987b) divisio nov.	Abbreviation of <i>L. positivus</i> positive, because four of the six classes stain Gram-positive	Acyl ester lipids; SRP with helices 1–4; SRP RNA lacks helix 6; lacking SRP 19p; ancestrally with murein; thick-walled (Gram-positive) or thin walled (Gram-negative); flagella with acid-soluble flagellin shafts; proteins with cleavable signal peptides secreted co-translationally via SRP or post-translationally via SecA; lacking <i>N</i> -linked glycoproteins – i.e. the traditional Firmicutes plus Mollicutes and Togobacteria	Order Bacillales
Subdivision 1. Endobacteria (Cavalier-Smith, 1998) subdivisio nov.	Gr. <i>endo</i> within, because spores are formed within an enveloping forespore cell	Low G+C content; without proteasomes; ancestrally with endospores	Order Bacillales

Table 1 (cont.)

Taxon	Etymology	Description	Type
Class 1. Togobacteria (Cavalier-Smith, 1992a) classis nov.	L. <i>toga</i> a loose outer garment, referring to the sometimes loose outer S-layer	Teichoic acid absent, stain Gram-negative; peptidoglycan layer thin; with a thin outermost S-layer or toga easily confused with the negibacterial outer membrane except at very high resolution; with or without endospores; heterotrophs not assigned to orders, e.g. <i>Selenomonas</i> , <i>Sporomonas</i> , <i>Dicryoglossus</i> , <i>Thermoanaerobibrio</i> , <i>Carboxydobacterium</i> ; anaerobic photoheterotrophs with bacteriochlorophyll <i>g</i> ;	Order Thermotogales
Class 2. Teichobacteria (Cavalier-Smith, 1998) classis nov.	Gr. <i>teichos</i> wall, because their walls contain teichoic acids	Heliobacteriales, e.g. <i>Heliobacteris</i> , <i>Heliobacterium</i> , <i>Heliophilum</i> ; and hyperthermophiles with acyl ether lipids, e.g. <i>Thermotoga</i> , <i>Petrogloia</i> , <i>Fervidobacterium</i>	Order Bacillales
Class 3. Mollicutes Edward and Freundt 1967		Thick rigid murein walls containing teichoic acids and lipoteichoic acid, stain Gram-positive; often form endospores; anaerobic or aerobic organoheterotrophs, e.g. <i>Bacillus</i> , <i>Streptococcus</i> , <i>Staphylococcus</i> , <i>Clostridium</i>	No type given
Subdivision 2. Actinobacteria* (ex Margulis 1974 as class) subdivisio nov.	Gr. <i>actino</i> ray, because of their often filamentous character and inclusion of all actinomycetes	Mycoplasmas; no endospores, peptidoglycan or teichoic acids, e.g. <i>Ureaplasma</i> , <i>Acholeplasma</i>	Order Actinomycetales
Class 1. Arthrobacteria* classis nov.	Gr. <i>arthron</i> joint, because of their often snapping division and the inclusion of the genus <i>Arthrobacter</i>	High G+C content with proteasomes; spores if present usually exospores; often with mycothiol instead of glutathione; predominantly aerobic; often with snapping division or branching filaments; phosphatidylinositol a major lipid; Gram-positive	Order Actinomycetales
Class 2. Arabobacteria classis nov.	N.L. <i>arabo-</i> combining form of arabic, because, unlike other bacteria, their cells or walls always contain arabinose, isolated originally from gum arabic	Cell walls varied, usually lacking diaminopimelic acid, usually with ornithine and/or lysine, never with arabinose; usually non-filamentous, lacking mycothiol or sterols, often facultative anaerobes; ancestrally with two layered walls and snapping division (e.g. <i>Arthrobacter</i> , <i>Actinomycetes</i> , <i>Propionibacterium</i> , <i>Bifidobacterium</i>)	Order Mycobacteriales
Order 1. Actinoplanales ord. nov.	From the genus <i>Actinoplanes</i>	Cell walls with meso-diaminopimelic acid, either glycine or arabinose and either galactose or xylose; non-filamentous cells sometimes with snapping division, e.g. <i>Corynebacterium</i> , fragmenting filaments, e.g. <i>Nocardia</i> , or branched filaments lacking aerial hyphae e.g. <i>Actinoplanes</i> ; frequently with mycolic acid, mycothiol and lipid-rich walls; some make cholesterol (<i>Mycobacterium</i>); commonly have phosphatidylethanolamine	Genus <i>Actinoplanes</i>
Order 2. Mycobacteriales ord. nov.	From the genus <i>Mycobacterium</i>	Walls with glycine not arabinose	Genus <i>Mycobacterium</i>
Class 3. Streptomycetes classis nov.	From <i>Streptomyces</i> , the best-known members	Typically with differentiated aerial filaments and spores; cell walls with meso or L-diaminopimelic acid, but no arabinose, galactose or xylose; aerobes with mycothiol, e.g. <i>Streptomyces</i> , <i>Frankia</i> ; lack phosphatidylethanolamine	Order Streptomycetales
Order 1. Streptomycetales ord. nov.	As for class above	As for class above	Genus <i>Streptomyces</i>
Division 2. Archaeobacteria (Woese & Fox, 1977) divisio nov.	Gr. <i>archae-</i> ancient	Syn. Mendosicutes (Gibbons & Murray, 1978); Metabacteria (Hori <i>et al.</i> , 1982); prenyl ether membrane lipids; signal recognition particles with 7S SRP RNA having helix 6 that binds SRP 19p, used both for membrane protein insertion and for all protein secretion; murein peptidoglycan, SecA and Hsp90 absent; co-translationally synthesized N-linked glycoproteins; flagellar shafts of acid-stable glycoprotein	Order Methanococcales

Table 1 (cont.)

Taxon	Etymology	Description	Type
Subdivision 1. Euryarchaeota (Woese <i>et al.</i> , 1990; rank Cavalier-Smith, 1998) subdivisio nov.	Gr. <i>eury</i> - broad; Gr. <i>archae</i> - ancient, because they have a wide range of archaeobacterial phenotypes	Ancestrally with core histones; cell walls varied; with FtsZ and eukaryote-like oligosaccharyl transferase	Order Methanococcales
Superclass 1. Neobacteria superclassis nov.	Gr. <i>neo</i> new; Gr. <i>bakterion</i> rod	Largest RNA polymerase subunit B split into two proteins; predominantly mesophiles; cell walls varied; with histones	Order Methanococcales
Class 1. Methanothermea* classis nov.	N.L. <i>methano</i> - combining form of methane; Gr. <i>therme</i> heat, because they all generate methane and some are hyperthermophiles	Methanogens with walls of pseudomurein (Methanobacteriales) or protein (Methanomicrobiales, Methanococcales, Methanopyrales); lacking DNA gyrase; ancestrally with reverse gyrase; sometimes hyperthermophiles, usually mesophiles	Order Methanococcales
Class 2. Archaeoglobae classis nov.	From the Archaeoglobales, the only member	Sulphate or nitrate reducing hyperthermophiles with glycoprotein walls; with tetraether lipids, DNA gyrase and reverse gyrase; sole order Archaeoglobales	Order Archaeoglobales
Class 3. Halomebacteria (Cavalier-Smith, 1986) classis nov.	Gr. <i>hals</i> salt; me- common scientific abbreviation for methane, since the class comprises both halophiles and somewhat halophilic methanogens	Biether lipids, DNA gyrase; often with complex carbohydrate walls; lack reverse gyrase; mesophilic methanogens, Methanosarcinales, uncultured marine euryarchaeotes and halobacteria, Halobacteriales	Order Halobacteriales
Superclass 2. Eurythermea* superclassis nov.	Gr. <i>eury</i> - broad; Gr. <i>therme</i> heat, because they are euryarchaeotes that are mostly hyperthermophilic or thermophilic	Ancestrally with cell walls of glycoprotein or protein; largest RNA polymerase subunit (B) unsplit; tetraether lipids	Order Thermococcales
Class 1. Protoarchaea classis nov.	Gr. <i>proto</i> first; Gr. <i>archae</i> - ancient, because they have all retained the putatively ancestral archaeobacterial phenotype of hyperthermophily, histones and sulphur reduction	With histones, reverse gyrase and cell walls; lacking DNA gyrase; hyperthermophiles, e.g. <i>Pyrococcus</i> , <i>Palaeococcus</i>	Order Thermococcales
Class 2. Picrophileae classis nov.	From the genus <i>Picrophilus</i>	Hyperacidophiles; membrane glycolipids and DNA gyrase; lacking methanogenesis, histones, reverse gyrase; with cell wall (<i>Ferroplasma</i> , <i>Picrophilus</i>) or surface coat (<i>Thermoplasma</i>); sometimes thermophiles	Order Picrophilales
Order 1. Picrophilales ord. nov.	From the genus <i>Picrophilus</i>	As for class above	Genus <i>Picrophilus</i>
Subdivision 2. Crenarchaeota (Woese <i>et al.</i> , 1990; syn. eocytes Lake) subdivisio nov.	Gr. <i>kren</i> spring, fount; Gr. <i>archae</i> - ancient	Sulphur-reducing respiration; with glycoprotein or protein cell walls, reverse gyrase and tetraether lipids; lacking FtsZ, eukaryote-like oligosaccharyl transferase and histones	Order Thermoproteales
Class 1. Crenarchaeota classis nov.	As for subdivision above	As for subdivision above. Thermoproteales, Sulfolobales – highly acidophilic, Desulfurococcales; cultured strains all hyperthermophiles	Order Thermoproteales
Cenarchaeales ord. nov.	Gr. <i>kainos</i> recent; Gr. <i>archae</i> - ancient, because their non-thermophily is a derived condition for archaeobacteria and they include <i>Cenarchaeum</i>	Mesophilic or psychrophilic crenarchaeotes	Genus <i>Cenarchaeum</i> Preston <i>et al.</i> 1996
Archaeobacteria <i>incertae sedis</i> : candidate group (possibly a crenarchaeote order) Korarchaeota (Barns <i>et al.</i> , 1996), recently cultured hyperthermophiles that tend to branch more deeply on 16S rRNA trees than others			

* Probably paraphyletic. The widespread dogma against paraphyletic taxa is misconceived and harmful (see Cavalier-Smith, 1998). The kingdom Bacteria is itself probably paraphyletic.

Table 2. Major archaeobacterial properties not found in eubacteria

(a) Neomuran properties (i.e. those shared with eukaryotes)
<ol style="list-style-type: none"> 1. Signal recognition particle (SRP) with 7S RNA with a helix 6 that binds SRP19 protein; protein secretion generally co-translational; SecA absent 2. Co-translational glycosylation of surface glycoproteins by transfer of GlcNAc and mannose-containing oligosaccharides from a dolichol isoprenoid carrier to <i>N</i>-asparagine; homologous oligosaccharyl transferases; murein absent 3. Ribosomal rRNA pseudouridylated by C/D-box snoRNAs 4. Core histones with histone fold [secondarily lost in some archaeobacteria (e.g. <i>Thermoplasma</i>) and some eukaryotes (dinoflagellates)] 5. Replicative DNA polymerases B type; inhibited by aphidicolin; replicative sliding clamp is PCNA-type, not part of a type C DNA polymerase holoenzyme; novel replication factor complex 6. Flap endonuclease and RAD2 DNA-repair enzymes 7. Seven or more RNA polymerase holoenzyme subunits (not four as in eubacteria) 8. Many similarities of ribosomal RNA and proteins; a more substantial projecting bill on the small ribosomal subunit; ribosomes insensitive to chloramphenicol; anisomycin inhibits peptidyl transferase by binding to 23S/28S rRNA 9. CCT-type group II chaperonins with eightfold symmetry, not sevenfold symmetry as in their distant eubacterial Hsp60 relatives; with built-in cap; co-chaperonin Hsp10 absent; prefoldin (GimC) channels nascent proteins to the chaperonin lumen 10. Some similar tRNA modification 11. Exosomes; complex of 11–16 proteins involved in exonucleolytic digestion of RNA; exonucleases, helicases and RNA-binding proteins (Koonin <i>et al.</i>, 2001) 12. More similar protein synthesis elongation factors (e.g. sensitive to ADP ribosylation by diphtheria toxin) 13. Co-translational selenocysteine insertion requires a SECIS-binding protein in addition to a selenocysteine-specific elongation factor 14. CCA 3' terminus of tRNA added post-translationally, not encoded by the gene 15. Protein synthesis initiated by methionine not <i>N</i>-formyl methionine; several extra initiation factors (eIF-2, 2A, 2B and 5A) 16. 5'-OH/3'-phosphate protein-spliced tRNA introns with homologous endonucleases 17. Novel type II DNA topoisomerase VI/meiotic protein 18. Insertion in catalytic subunit of the vacuolar-type proton-pumping ATPase 19. Hexameric replicative DNA helicase Mcm instead of eubacterial DnaB (Poplawski <i>et al.</i>, 2001)
(b) Unique archaeobacterial properties
<ol style="list-style-type: none"> 1. Prenyl ether instead of acyl ester lipids 2. Flagellar shaft of acid-insoluble glycoproteins related to pilin, not acid-soluble flagellin 3. DNA-binding protein 10b 4. Unique tRNA modifications, including archaeosine in D-loop and absence of queuine 5. A tiny large subunit ribosomal protein, LX 6. Absence of Hsp90 chaperone 7. RNA polymerase A split into two proteins 8. Glutamate synthetase split into three separate proteins

eubacteria. Although the differences in organization of the replication, transcription and translation machinery of archaeobacteria are well known (Doolittle, 1998; Graham *et al.*, 2000), the full extent of other major differences between archaeobacteria and eubacteria in cell organization is still insufficiently widely appreciated, some having only become apparent recently. Table 2 lists the key differences between archaeobacteria and eubacteria. The scale of these is so great that this paper, which attempts to explain them all, is necessarily long and detailed. To help the reader see the wood for the trees, let me outline its basic structure. I shall argue

that all 19 features listed in Table 2(a) arose in the common ancestor of eukaryotes and archaeobacteria in association with the loss of eubacterial peptidoglycan and its functional replacement by neomuran *N*-linked glycoproteins. This part of the neomuran theory is identical to the original, except that the number of uniquely shared neomuran character suites has doubled since the theory was originally proposed (Cavalier-Smith, 1987b), placing the relationship between eukaryotes and archaeobacteria beyond question. To save space, I refer readers to the original paper for more details of the basic rationale of the neomuran

theory, including the sister relationship of archaeobacteria and eukaryotes (rather than an ancestor descendant one, as suggested by Van Valen & Maiorana, 1980; Rivera & Lake, 1992; Baldauf *et al.*, 1996), the much more ancient ancestral character of eubacteria and the changeover from peptidoglycan to glycoproteins, as well as for diagrams summarizing the cellular transformations (Cavalier-Smith, 1987b).

I concentrate here on six things. First are the key innovations of the present paper: the arguments that the majority of the novel neomuran characters arose as adaptations of the neomuran ancestor to thermophily and that nearly all neomuran characters can be used to polarize unambiguously the direction of evolution from posibacteria to neomura, not the reverse. Second is the argument that, after the neomuran common ancestor adapted thus to thermophily, the archaeobacterial ancestor alone underwent a more extreme adaptation to hyperthermophily and hyperacidity that produced almost all the uniquely archaeobacterial characters listed in Table 2(b). About a third of the paper discusses the origin of each of these neomuran and archaeobacterial characters. Having provided extensive evidence from comparative biology that neomura are derived compared with eubacteria, I then discuss the fossil record for all three domains of life, which shows exactly the same thing and indicates that neomura are about four times younger than eubacteria. Central to my re-evaluation of the fossil record is recent evidence that some actinobacteria, the probable ancestors of eukaryotes, make sterols (Lamb *et al.*, 1998), which invalidates earlier palaeontological interpretations of fossil steranes as eukaryotic markers; this and other recent discoveries of morphological fossils make my earlier estimate of 850 My for the origin of eukaryotes (Cavalier-Smith, 1980) more accurate than more recent ones giving an older date (Cavalier-Smith, 1987a, 1990). My fourth topic is the application of the ideas of quantum and mosaic evolution to the interpretation of molecular sequence trees. These principles explain many of the puzzling conflicts between different trees. Still more importantly, in conjunction with my discussion of the evidence for temporarily accelerated evolution affecting all the characters of Table 2 at the time of origin of neomura, but not other more ancestral characters, they tell us that reciprocally rooted protein paralogue trees and single-gene trees (e.g. for rRNA) based on them are so dimensionally distorted as to be highly misleading about the temporal history of life; this has caused the misrooting of the universal tree of life. Once we understand these distortions, we can see that there is no genuine conflict between any molecular trees and the fossil evidence that neomura are very recent. My fifth topic is to use this new understanding of the strengths and weaknesses of different molecular trees to integrate their evidence with the fossil record and cell-biological considerations so as to pinpoint the root of the tree as accurately as is currently possible. I shall argue that recent evidence concerning the evol-

ution of photosynthesis strongly supports earlier arguments that the root of the tree of life lies within the negibacteria (Cavalier-Smith, 1987a, b, 1991a, b, 1992b). Although the precise position of the root remains uncertain, it very likely lies within or immediately adjacent to the green bacteria, as suggested previously (Cavalier-Smith, 1985a, 1987a). I point out that many current interpretations of cell and molecular evolution are fundamentally flawed by the serious misrooting of molecular trees and the misplacing of some long branches. My sixth concern is to show that, although lateral gene transfer is more frequent and confusing in bacteria than in eukaryotes, we can still construct sensible organismal phylogenies for bacteria, provided we emphasize organismal features that depend on strong co-adaptation between macromolecules and do not overemphasize the evidence from any single molecule.

I emphasize that, for most of the history of life, immensely long periods of relative stasis have followed two explosive radiations or 'biological big bangs', each stimulated by revolutionary innovations in cell biology: (i) the origin about 3700 My ago of the first eubacterial cell with peptidoglycan walls and photosynthesis (Cavalier-Smith, 2001) and (ii) the origin about 850 My ago of the ancestral neomuran cell, when *N*-linked glycoproteins replaced peptidoglycan and the pre-eukaryote neomurans evolved phagotrophy, internal skeletons and the endomembrane system. The neomuran theory of the origin of eukaryotes is further developed in another paper, published separately because of space constraints (Cavalier-Smith, 2002); however, the two papers need to be read together fully to appreciate and evaluate this revised neomuran theory of the simultaneous actinobacterial origins of archaeobacteria and eukaryotes. A third paper, on the origin of the negibacterial cell and the genetic code (Cavalier-Smith, 2001), is complementary to both, since it shows that it is much easier to understand the origin of life if we root the tree among photosynthetic negibacteria, rather than between archaeobacteria and eubacteria as suggested by most reciprocally rooted protein paralogue trees.

I also discuss the early diversification of negibacteria that constituted the first big bang, integrating both fossil and recent evidence and arguing that the differences between the six phyla arose primarily as divergent adaptations within the microlayers of early microbial mats. In addition to these phylogenetic and evolutionary questions, I discuss briefly the higher classification of bacteria and how it may be improved.

Secondary hyperthermophily and acidophily and the origin of archaeobacteria

It has long been argued that the prenyl ether lipids of archaeobacteria evolved as replacements for the acyl ester lipids of eubacteria (Cavalier-Smith, 1987a, b) as a secondary adaptation to hot, acid environments

Fig. 1. The origin and diversification of Archaeobacteria. Archaeobacteria originated by two successive revolutions in cell biology: a neomuran phase shared with their eukaryote sisters followed shortly by a uniquely archaeobacterial one. The first, neomuran phase was an adaptation to thermophily and involved a really major transformation of 19 key characters, including replacement of the cell wall peptidoglycan murein by *N*-linked glycoprotein and a great upheaval in the cell's protein-secretion and DNA-handling machinery. The second, relatively minor phase of specifically archaeobacterial innovations, notably replacement of acyl ester membrane by isoprenoid tetraether lipids and of eubacterial flagellin by glycoproteins, involved further adaptations to hyperthermophily and hyperacidity, respectively. Substantially later, several lineages independently readapted secondarily to mesophily. Lateral transfer of genes from the immensely older and far more diverse eubacteria often played a role in these secondary returns to mesophily and may also have done in the origins of archaeobacterial hyperthermophily, sulphate reduction by *Archaeoglobus* and methanogenesis. This phylogenetic interpretation is based on a synthesis of discrete organismal and molecular characters treated cladistically, sequence trees and palaeontology, as discussed in the text.

(Reysenbach & Cady, 2001). The presence of sulphur-dependent hyperthermophiles among both euryarchaeotes (*Thermococcales* and *Methanothermus*) and crenarchaeotes strongly suggests that the ancestral archaeobacterium was also a sulphur-dependent hyperthermophile (Woese, 1987; Barns *et al.*, 1996). It is very unlikely, however, that the ancestral eubacterium or first cell was a thermophile or hyperthermophile, as is sometimes suggested (Achenbach-Richter *et al.*, 1987; Pace, 1991); the low thermal stability of essential organic molecules such as RNA makes it far more likely that the first cell was a mesophile (Levy & Miller, 1998) or even psychrophile (Cavalier-Smith, 2001). Hyperthermophilic environments were probably the last to be colonized; the chimaeric origin of reverse gyrase implies that hyperthermophiles evolved last of all (Forterre, 1996) and maximum-likelihood reconstruction of the cenozoic base composition favours a mesophile (Galtier *et al.*, 1999). The distribution of reverse gyrase within archaeobacteria indicates that it was present in their common ancestor but was lost by Halobacteria and

Methanosarcinales (here grouped together as Halomebacteria; Table 1) and by *Thermoplasma* (López-García, 1999) and replaced by eubacterial DNA gyrase by lateral gene transfer. The fact that mesophilic methanogens and halobacteria share a split RNA polymerase gene (RpoB protein exists as two distinct subunits) uniquely with Archaeoglobales (Klenk *et al.*, 1997) implies strongly that this clade (which I call Neobacteria; Table 1), and thus the mesophily of Halomebacteria, is derived within the Archaeobacteria (Fig. 1) and that reverse gyrase was replaced by DNA gyrase independently in the thermophile *Thermoplasma*, which has the ancestral unsplit RNA polymerase gene. This secondary mesophily of Halomebacteria was associated with the replacement of tetraether prenyl lipids, which form thermostable monolayers, by biether prenyl lipids giving more fluid bilayers. I argued previously (Cavalier-Smith, 1987a) that there would probably be no selective advantage for a secondary mesophile in replacing these lipids by eubacterial/eukaryotic acyl ester lipids, whereas replacement of acyl esters by the more heat-stable and

acid-stable prenyl ethers (initially tetraethers in the archaeobacterial ancestor), which are much more impermeable to protons at higher temperature and clearly adaptive to hot acid (Albers *et al.*, 2000), would undoubtedly be selectively advantageous to a hyperthermophile that evolved secondarily from a mesophilic ancestor. Thus, both the unique lipids and reverse gyrase indicate strongly that the direction of evolution was from mesophilic eubacteria to hyperthermophilic archaeobacteria. Because of their novel lipids, thermophilic archaeobacteria can control their pH and use proton gradients as energy sources, unlike eubacterial thermophiles (Albers *et al.*, 2000).

I now argue that secondary acidophily gives a simple adaptive explanation to the otherwise puzzling fact that archaeobacterial flagellar shafts lack classical flagellins but are built of unrelated proteins (Faguy *et al.*, 1994). Eubacterial flagellin filaments disassemble to monomers under very acid conditions, and at somewhat acid pH eubacterial flagella undergo a remarkable phase transition to an abnormal, less efficient, curly form (Kamiya *et al.*, 1982). Replacing an ancestral flagellin polymer by recruiting an acid-stable glycoprotein from pili, which the shaft resembles (Faguy *et al.*, 1994), would have enabled archaeobacterial flagella to function in highly acid conditions, while retaining the same basal rotary motor. As archaeobacterial flagella operate well in neutral conditions, there would be no selective advantage in replacing them by flagellin in secondary mesophiles.

Paucity of unique features of archaeobacteria

Apart from the special lipids and flagellar shafts, only four other unique features have been so far identified as generally present in archaeobacteria (Table 2b). Best known are the unique post-transcriptional modifications of their tRNAs. These are also almost certainly secondary adaptations to thermophily. Kowalak *et al.* (1994) have shown that modifications greatly increase the thermal stability of archaeobacterial tRNAs and are more extensive at higher temperatures. Unlike any other organisms, archaeobacteria replace a guanine at position 15 in the D-loop with archaeosine. Unlike eubacteria and eukaryotes, they never use queuine in the wobble position of the anticodon. Graham *et al.* (2000) assert that the transglycosylase protein that inserts archaeosine is unique to archaeobacteria. This is only half true. The enzyme is clearly homologous in sequence to the queuine-inserting one of other organisms. Rather than being a novel archaeobacterial enzyme, a pre-existing one simply switched its specificity; I suggest that the switch was from queuine insertion to archaeosine insertion. A more convincingly unique archaeobacterial protein is the small, 10 kDa DNA-binding protein, 10b; in *Sulfolobus*, its binding produces passive negative supercoils at very high temperatures, but not at mesic ones (Xue *et al.*, 2000). I suggest that it evolved as a secondary adaptation to hyperthermophily in the ancestral archaeobacterium, as a substitute for active negative

supercoiling by DNA gyrase, which was lost in the neomuran ancestor (i.e. their last common ancestor; Fitch & Markowitz, 1970). Too little is known about the other unique archaeobacterial protein, the tiny, 77-amino-acid protein LX of the large ribosomal subunit, to know whether it was also an adaptation to hyperthermophily or hyperacidity or evolved for another reason. There is no reason to think that the splitting of the RNA polymerase gene A to make two separate proteins or of glutamate synthetase into three were adaptations; both were possibly neutral changes that became incidentally fixed in the archaeobacterial ancestor.

Graham *et al.* (2000) identified 36 conserved hypothetical proteins found in all five of the archaeobacterial genomes then sequenced. When their functions are known, it will be interesting to see how many are also adaptations to hyperthermophily and how few are really unique to archaeobacteria. It is likely that most are evolutionarily related to eubacterial proteins, but diverged so drastically during archaeobacterial origins that a relationship is not obvious from the sequences. Adaptation to hyperthermophily increases the charged residues in proteins (Cambillau & Claverie, 2000); in some proteins, such adaptation may have led to much more extensive changes. Four proteins claimed to be unique to archaeobacteria (Graham *et al.*, 2000) clearly are not. Both A and B subunits of DNA topoisomerase VI, though stated to be uniquely archaeobacterial [Forterre & Philippe, 1999; Graham *et al.*, 2000 (mislabelled as topoisomerase IV in their table)], are strongly related to those of a meiosis-specific protein of eukaryotes. As no eubacterial relatives are certainly known, Table 2 shows them as neomuran, not simply archaeobacterial characters; however, I shall argue that this neomuran topoisomerase evolved from DNA gyrase and is not a novel protein. The Holliday junction cleavage resolvase is not unique to archaeobacteria (Graham *et al.*, 2000) but is structurally related to other nucleases widespread in eubacteria (Aravind *et al.*, 2000); it even has distant primary structure similarity to RuvC, the eubacterial Holliday junction resolvase, despite the latter's different fold. The transcription termination/inhibition factor is clearly related to the functionally equivalent NusG protein of eubacteria and to elongation factor Spt5 of eukaryotes.

To speak of an archaeobacterial genomic signature (Graham *et al.*, 2000) is misleading. The genome organization of archaeobacteria is fundamentally the same as that of eubacteria. What is unique are the rather small number of genes mentioned above, and even their uniqueness is probably exaggerated by accelerated sequence evolution. Graham *et al.* (2000) also exaggerate the uniqueness of archaeobacteria by referring to features present in only some archaeobacteria as 'archaeobacterial signatures'. However, such features as methanogenesis are not properties of archaeobacteria as a whole; it is as misleading to call them archaeobacterial signatures as it would be to call

feathers or hair ‘vertebrate signatures’, rather than bird or mammal signatures. Many hundreds of proteins they call archaeobacterial signatures are actually euryarchaeote, crenarchaeote or methanogen signatures, for example, and are irrelevant to the understanding of the origin of archaeobacteria, my main focus. However, some features found only in archaeobacteria, but not in all, will have been present in their cenancestor but lost by a few lineages. The most obvious of these are the two flagellar shaft proteins and a flagellar accessory protein, which Table 2 does treat as general archaeobacterial proteins. Phylogenetic analysis will eventually reveal other non-universal archaeobacterial proteins that were actually also cenancestral.

Thus, all functionally understood unique and general features of archaeobacteria are apparently adaptations to hyperthermophily or hyperacidity. There is no reason to think that any are ancient or relics of early evolution. Archaeobacteria are genomically and cytologically fundamentally the same as posibacterial eubacteria. Their uniqueness among bacteria rests much less on the small number of unique archaeobacterial characters (Table 2b) than on the very large number of characters shared with eukaryotes (Table 2a). These are especially important, as many are not single-gene characters but depend on numerous genes. Thus, it is exceedingly misleading to refer to archaeobacteria as a third form of life. Except for their membrane lipids, flagellar shafts, tRNA modifications and the small proteins 10b and LX, they share virtually every understood character with other bacteria or with their eukaryote sisters. Because the origin of neomuran characters (Table 2a) is important for understanding the origin of both archaeobacteria and eukaryotes (Cavalier-Smith, 1987b, 2002), I discuss them first.

Novel cell walls, thermophily and the origin of neomura: rooting the tree in Eubacteria

At first sight, the changes listed in Table 2(a) seem an arbitrary set of molecular properties from the thousands that characterize bacteria. The neomuran theory argues, however, that virtually all are explicable as co-ordinated changes in the cell envelope and interactions of ribosomes with it or with the adaptation of chromatin to thermophily. None involves changes in intermediary metabolism, to which the majority of eubacterial genes are devoted. It is the interconnections between so many of these changes that are evolutionarily important. The replacement of peptidoglycan by glycoprotein involved novel protein-secretion mechanisms; these involved changes in the ribosomes and in the chaperone machinery. Changes in DNA-binding proteins affected the replication, repair and transcription machinery. Though I shall discuss them one by one, the key point of the theory lies in the concerted evolution that radically transformed hundreds of genes and their interacting gene

products during one short evolutionary episode, but left thousands more little changed.

The first 11 neomuran characters (Table 2a) can be interpreted as adaptations by the ancestral neomuran to thermophily; as they have generally not been reversed either in secondarily mesophilic euryarchaeotes or in eukaryotes, which probably became mesophiles during eukaryogenesis (Cavalier-Smith, 2002), this indicates that the reverse change from the neomuran to the eubacterial state would be unlikely to be positively selected. Thus, these 11 characters plus the first four unique archaeobacterial characters, which have also not been reversed in secondary mesophiles, together provide 15 evolutionary ‘valves’ that we can use with high confidence to polarize the direction of evolution from eubacteria to neomura rather than the reverse. Elsewhere (Cavalier-Smith, 2001), I argued that characters 12–14 are so much more complex than those of eubacteria that each must be regarded as derived not primitive, while tRNA introns must be derived (Cavalier-Smith, 1991c). The insertion in the catalytic subunit of the vacuolar proton-pumping ATPase, though an apparently trivial character unconnected with the others, is important because its absence in the paralogous non-catalytic subunit strongly indicates that the eubacterial condition is ancestral and the neomuran one was derived by an insertion in the common ancestor of eukaryotes and archaeobacteria (Gogarten & Kibak, 1992). The selective advantage of this universally conserved change probably lies in the increased complexity of the linker proteins that join the ATPase to the membrane-spanning proteolipid that also underwent great change in the neomuran cenancestor (Hilario & Gogarten, 1998). Overall, therefore, including reverse gyrase, there are 20 different characters, most rather complex, that independently polarize the direction of evolution from eubacteria to neomura. Not one supports the reverse. Fig. 2 summarizes this view of the tree of life.

First, consider the switch in protein-secretion mechanism between eubacteria and neomura.

Derived neomuran protein-secretion and -glycosylation mechanisms

In eubacteria, proteins bearing a signal sequence follow two distinct pathways. Membrane proteins with an uncleaved signal sequence are inserted co-translationally directly into the cytoplasmic membrane by the interaction of their signal sequence with the signal recognition particle (SRP); this causes ribosomes to dock onto a ribosome receptor (the SecYEG protein complex) embedded in the membrane. Secretory proteins with cleavable signal sequences, by contrast, are often released from the ribosome into the cytosol and are recognized by SecA protein, which directs them post-translationally to the SecYEG channel for translocation across the membrane into the periplasmic space. In proteobacteria like *Escherichia coli*, most, if

Fig. 2. The rooted tree of life, showing key innovations. The ancestral eubacterial domain is about four times older than the archaeobacteria and eukaryotes, which jointly form a recent clade, designated neomura (Cavalier-Smith, 1987b) because the ancestral eubacterial peptidoglycan was replaced by *N*-linked glycoprotein during their common origin about 850 My ago. Both the fossil record and the 20 character suites (Table 2) that polarize the tree from eubacteria to neomura prove that eubacteria are ancestral and paraphyletic – the only Ur-domain. The double envelope of negibacterial cells probably evolved well before the cenancestor by the fusion of obcells, as described elsewhere (Cavalier-Smith, 2001); it was retained as the double envelope of mitochondria and chloroplasts when they originated from proteobacteria and cyanobacteria. The negibacterial outer membrane was lost only once in the history of life, in the ancestral posibacterium. This unimembranous character of posibacteria was a pre-adaptation for the much later origin of neomura from a thermophilic actinobacterium similar to a mycobacterium. After the origin of the 19 shared character suites (Table 2a), the neomuran ancestor diverged sharply into two contrasting lineages; one formed a glycoprotein wall and became hyperthermophilic, evolving prenyl ether lipids and losing many eubacterial genes, e.g. for H1 histones, to form the archaeobacteria, the other became much more radically changed by using its glycoproteins as a flexible surface coat, evolving phagotrophy, an endomembrane system, endoskeleton and nucleus (N) and enslaving an α -proteobacterium as a protomitochondrion (M) to become the first eukaryote, as explained in detail elsewhere (Cavalier-Smith, 2002).

not all, secretory proteins follow this post-translational pathway. In *Bacillus subtilis* (and, very likely, other posibacteria), however, only a minority of secretory enzymes use the SecA post-translational mechanism; the great majority are probably secreted co-translationally by the SRP mechanism (Tjalsma *et al.*, 2000).

Neomura, however, do not have SecA and secrete essentially all proteins with a cleavable signal sequence co-translationally. This is achieved through the presence of an additional translation-arrest domain on the SRP RNA and an extra 19 kDa SRP protein. The translation-arrest domain delays the extension of the polypeptide chain sufficiently for the signal sequence to bind to the membrane receptor and for translation across the membrane to be initiated prior to the cleavage of the signal peptide by the membrane-associated signal peptidase, which has its active site on the periplasmic surface of the cytoplasmic membrane (Mason *et al.*, 2000; Walter *et al.*, 2000). This direct co-translational threading of the nascent polypeptide

through the cytoplasmic membrane to the outside, where it can fold immediately into its native configuration, would be especially advantageous for a thermophile. With the eubacterial post-translational SecA-based system, there is a much greater risk that the protein could become irreversibly denatured in the cytosol and lose its translocation competence or be degraded by cytosolic proteases that recognize unfolded proteins. Eubacteria possess two other purely post-translational translocases, TAT and YidC (Stuart & Neupert, 2000; Samuelson *et al.*, 2000). These latter systems would also be more prone to disruption by heat, which might denature proteins irreversibly before they ever reached the membrane, than would an obligately co-translational one.

The ancestral bacterium is unlikely to have done much protein secretion compared with the more complex modern ones, and the smallest, simplest SRP of negibacteria is likely to be the ancestral type that evolved initially just for the insertion of membrane proteins, essential even for the simplest, most primitive cell

Table 3. Neomuran characters shared by some or all actinobacteria but not other eubacteria

The ability to produce *N*-penicillin and cephalosporins is shared, as far as is known, only by fungi, which may therefore have acquired them by lateral gene transfer. The listed, more generally distributed eukaryotic characters are more likely to have been inherited vertically.

General neomuran characters
<ol style="list-style-type: none"> 1. Proteasomes 2. 3'-Terminal CCA of tRNAs mostly (actinobacteria) or entirely (neomura) added post-transcriptionally
Characters shared by eukaryotes generally but not archaeobacteria
<ol style="list-style-type: none"> 1. Sterols 2. Chitin 3. Numerous serine/threonine phosphotransferases and protein kinases related to cyclin-dependent kinases (Av-Gay & Everett, 2000) 4. Tyrosine kinases 5. Long H1 linker histone homologues related to eukaryotes ones throughout 6. Calmodulin-like proteins (Swan <i>et al.</i>, 1987)* 7. Phosphatidylinositol (in all actinobacteria) 8. Three-dimensional structure of serine proteases 9. Primary structure of alpha amylases 10. Fatty acid synthetase a complex assembly 11. Desiccation-resistant exospores 12. Double-stranded DNA repair Ku protein with C-terminal HEH domain (Aravind & Koonin, 2001)

* Xi *et al.* (2000) report a protein with calmodulin-like motifs, but its sequence is much less similar to calmodulin than those of Streptomyces and Arabobacteria, which are remarkably like those of sarcoplasmic reticulum.

(Cavalier-Smith, 2001). As mesophilic eubacteria became more complex and started to secrete proteins, they added the SecA mechanism to facilitate this. One particular phylum alone, the Proteobacteria, made the further addition of the SecB chaperone to reduce the problem of denaturation and degradation of proteins prior to secretion. I suggest that this problem became particularly acute in actinobacteria, which are often thermophiles (never hyperthermophiles) that secrete an unusually large number of proteins or peptides; pronounced protein secretion is a basic characteristic of posibacteria – *Bacillus subtilis* secretes over 300 (Tjalsma *et al.*, 2000); its more-complex SRP has helices 1–4 like neomura. These two features of their lifestyle may explain why the ancestral actinobacterium evolved proteasomes for the degradation of misfolded or denatured proteins. Proteasomes are constitutively synthesized, cylindrical macromolecular assemblies in which protein digestion takes place within the cylinder and which are found only in Actinobacteria and neomura (Maupin-Furlow *et al.*, 2001). This is one of a dozen important reasons (Table 3; discussed briefly later in this paper and in more detail by Cavalier-Smith, 2002) why Actinobacteria are the most likely ancestors of neomura. It used to be thought that *Thermoplasma*, in which archaeobacterial proteasomes were first characterized, also had ubiquitin, like eukaryotes, but the genome sequence contradicts this (Ruepp *et al.*, 2000). Since ubiquitin has not been convincingly demonstrated in any archae-

bacterium, but is universal in eukaryotes, I suggest that proteasome evolution occurred in two temporally distinct phases: first, the origin of the basic 20S proteasome in the common ancestor of actinobacteria and neomura, then, very much later, in the pre-eukaryotic lineage alone, the evolution of ubiquitin and the polyubiquitin system for tagging proteins for degradation that the more complex eukaryotic 26S proteasome uses. I argue elsewhere (Cavalier-Smith, 2002) that the eukaryotic complexification of the proteasome was connected with the evolution of novel eukaryotic cell-cycle controls. The basic 20S proteasome, though a prerequisite for these later elaborations, had quite other origins in an early actinobacterium. The fact that inhibition of proteasome action in *Thermoplasma* has much more severe effects during heat shock than in normal growth (Ruepp *et al.*, 1998) supports my thesis that proteasomes were initially an adaptation to thermophily.

The narrow openings of the proteasome nanocompartment (Maupin-Furlow *et al.*, 2001) would allow denatured proteins to enter, but not native ones, and probably not those complexed with SecA. All other eubacteria lack proteasomes but have an HslUV energy-dependent protease instead, which is inducible by heat shock. It is reasonable to regard HslUV, which mediates the heat-shock response of most eubacteria, as an adaptation by a mesophile to temporarily hot

conditions. I suggest that this was replaced by the constitutive proteasome in a thermophilic common ancestor of Actinobacteria (the G+C-rich posibacteria, the base composition of which is equally reasonably interpretable as a secondary adaptation to thermophily). Having a constitutive proteasome would, however, increase the risk of post-translationally secreted proteins becoming denatured and degraded before they could be secreted, especially if protection by being bound to SecA was only partial. Evolution of the SRP translation-arrest domain would solve this problem and also make SecA no longer useful, and thus lost rapidly in the ancestral neomuran.

Archaeobacterial *N*-linked glycoproteins are made co-translationally by the transfer of complex oligosaccharides by a membrane-bound oligosaccharyl transferase from a dolichol phosphate carrier to the nascent protein associated with membrane-bound ribosomes (Lechner *et al.*, 1986), a process absent in eubacteria. In both eubacteria and archaeobacteria, ribosomes are attached to the cytoplasmic membrane by an SRP (Walter *et al.*, 2000), in which Ffh protein recognizes the signal peptide and binds to the membrane SRP receptor FtsY, after which the protein is inserted by a trimeric SecYEG protein translocase. Eubacteria differ from archaeobacteria and eukaryotes, however, in having a smaller SRP with 4·5S RNA not 7S RNA, which is associated with their translocation being more often post-translational: many eubacterial proteins are brought to the SecYEG translocase post-translationally with the help of SecA protein and soluble SecB or other chaperones. However, in the ancestor of archaeobacteria and eukaryotes, the eubacterial 4·5S SRP RNA (Walter *et al.*, 2000) became extended to the neomuran 7S SRP by the addition of an elongation-arrest domain (Mason *et al.*, 2000); this prevents the secretory or membrane protein emerging until the ribosome binds to the membrane, thereby protecting it more simply than would a complex chaperone system from premature denaturation. Elongation arrest presumably allowed archaeobacteria to rely less on the eubacterial post-translational secretion mechanism and to dispense with Hsp90 chaperone and SecA and the TAT and YidC translocases: 'thermal streamlining'. But, if archaeobacterial 7S SRP had been the ancestral type, I cannot see why it should have been reduced to the 4·5S SRP, which would reduce the efficacy of a perfectly good co-translational system. A key factor in the greater emphasis on co-translational transfer may have been the origin of the co-translational *N*-linked glycosylation of novel wall proteins, a key pre-adaptation for the origin of eukaryotes (Cavalier-Smith, 1987b). Note that the relatively much greater role for the co-translational mechanism in Posibacteria compared with Proteobacteria means that they are partially pre-adapted for the evolution of the neomuran system. The relative importance of these two mechanisms in other eubacterial phyla is unknown, but ought to be studied.

Analogous arguments may account for the absence of the characteristically eubacterial Clp A protease activity from the neomuran cytosol (secondarily reacquired by eukaryotes in mitochondria and chloroplasts and modified into an ATPase in archaeobacteria).

The changeover from partially post-translational to essentially exclusively co-translational protein secretion discussed above was a key pre-adaptation for the evolution of the rough endoplasmic reticulum (RER) and, therefore, the entire eukaryotic endomembrane system, as discussed in detail separately (Cavalier-Smith, 2002).

The second fundamental neomuran innovation was the evolution of *N*-linked glycoproteins. I consider that the biosynthesis of *N*-linked glycoproteins is too complex to have been present in the first cell. A complex, mannose-rich oligosaccharide core attached to an isoprenoid carrier (dolichol phosphate) by two residues of *N*-acetylglucosamine (GlcNAc) is synthesized by a suite of different enzymes and moved to the non-cytosolic face of the membrane (RER in eukaryotes; cytoplasmic membrane in archaeobacteria; Zhu & Laine, 1996) by the highly hydrophobic dolichol. Core oligosaccharides are cleaved from the dolichol and ligated to asparagine residues of proteins partly translocated across the membrane by the SRP-associated machinery discussed above. Although such glycoproteins might, in principle, have evolved in eubacteria for some proteins, there is no evidence that they did so prior to the loss of SecA and the exclusive reliance on co-translational secretion. As general co-translational insertion is probably a direct adaptation to thermophily, the origins of *N*-linked glycoproteins may be regarded as an indirect consequence of thermophily. The immediate selective force, however, may have been the resistance it gave the early neomuran to the β -lactam and other antibiotics that inhibit murein biosynthesis or enzymes (lysozyme) that digest murein and which were secreted by its actinobacterial relatives. Thus, this innovation makes adaptive sense if it occurred in an environment such as soil and rotting organic matter, rich in posibacterial synthesizers of antibiotics. Glycoprotein is, in effect, an antibiotic-resistant replacement for murein peptidoglycan; as muramopeptides are the target of β -lactams, muramic acid (one of the two aminosugars constituting eubacterial peptidoglycan) was lost but the other aminosugar, GlcNAc, was retained as part of the core oligosaccharide of *N*-linked glycoproteins.

The above arguments make it easy to understand the changeover from eubacterial murein and partially post-translational protein secretion to neomuran glycoprotein and co-translational protein secretion. I can see no adaptive reason why either change should have gone in the reverse direction. The streamlined neomuran protein secretion is just as good for mesophiles as thermophiles. Neither mesophily nor the absence of β -lactam antibiotics in an environment would favour the replacement of glycoprotein by murein. Thus, these two changes, like the three discussed earlier,

unequivocally polarize evolution from eubacteria to archaeobacteria, not the reverse. Most rRNA and several protein trees indicate that methanogenesis evolved after the divergence between euryarchaeotes and crenarchaeotes. Interestingly, some methanogenic archaeobacteria (Methanobacteriales) have secondarily evolved pseudomurein as a replacement for glycoprotein (ability to make other glycoproteins is retained). Far from being a sign of antiquity (Stackebrandt & Woese, 1981), the novel pseudomurein is probably a late adaptation. GlcNAc is found in pseudomurein, as in glycoproteins, but, as the ancestral neomuran had entirely lost the capacity to make muramopeptides, no murein is present and the sugar structure is novel.

Thermophily and the origins of H1, core histones and DNA topoisomerase VI

I also interpret the origin of core histones as an adaptation to thermophily, to induce negative supercoiling passively by wrapping the DNA round proto-nucleosomes, with less exorbitant energy costs than using DNA gyrase. In itself, this does not tell us the direction of evolution, since core histones can be lost, as they have been within eukaryotes in peridinean dinoflagellates, where only H1 homologues remain (Kasinsky *et al.*, 2001), and in euryarchaeotes, where histones have been lost in *Thermoplasma* (Ruepp *et al.*, 2000). Histones have not been found in crenarchaeotes, which have eubacterial-type DNA-binding proteins. Although it is possible that core histones evolved only in the ancestral euryarchaeote and were never present in crenarchaeotes, that could be so only if eukaryotes evolved directly from euryarchaeotes (Sandman & Reeve, 1998). If, as most evidence and evolutionary arguments suggest, eukaryotes are sisters to archaeobacteria as a whole and not direct euryarchaeote descendants (Cavalier-Smith, 2002), core histones must have evolved in the ancestral neomuran and have been lost in stem crenarchaeotes (unless they moved between eukaryotes and euryarchaeotes by lateral gene transfer, a possibility that I acknowledge but strongly discount). Hyperthermophilic crenarchaeotes are even more extreme hyperthermophiles than euryarchaeote hyperthermophiles. Since the psychrophilic or mesophilic crenarchaeotes (Cenarchaeales) are phylogenetically derived (DeLong *et al.*, 1998), the ancestral crenarchaeote probably adapted to a hotter habitat than did any previous bacteria; I suggest that this caused the loss of histones and replacement of their function by other proteins, for example the 66 amino acid Sac7d DNA-binding protein of *Sulfolobus*, which is exceedingly heat- and acid-stable and sharply kinks and stabilizes DNA by intercalation (Robinson *et al.*, 1998).

Although core histones probably evolved in a stem neomuran, H1 linker histones did so much earlier, in the eubacterial ancestors of neomura (Kasinsky *et al.*, 2001). The fact that the H1 homologue of the actinomycete *Streptomyces* is more similar in length and

sequence to that of eukaryotes than any yet known from non-actinobacterial eubacteria (Kasinsky *et al.*, 2001) is another piece of evidence for the actinobacterial ancestry of neomura. The ancestral eukaryote retained both the actinobacterial H1 histone (many actinobacteria are thermophiles) and the novel neomuran core histones. By contrast, their sisters, the ancestral archaeobacteria, must have lost H1 and retained the novel core histone.

The ancestral neomuran also lost eubacterial DNA gyrase activity and evolved a novel type II DNA topoisomerase (topoisomerase VI), not found in eubacteria; in eukaryotes, the homologous enzyme is restricted to meiosis and makes the double-stranded breaks needed for crossing over, which probably evolved as their ancestor was readopting mesophily. The B subunit of topoisomerase VI is very distantly related to the B subunit of DNA gyrase, but the A subunit (Spo11 in eukaryotes) is much shorter than that of DNA gyrase. DNA gyrase can be converted artificially to a conventional type II topoisomerase by deleting the C-terminal region of the A subunit responsible for the active wrapping of DNA (Kampranis & Maxwell, 1996). I suggest that this happened naturally in the ancestral neomuran as a direct consequence of the evolution for the first time of passive negative supercoiling by histones; this made active negative supercoiling by DNA gyrase redundant, so mutational truncation of the GyrA subunit was no longer disadvantageous and the former gyrase evolved rapidly into the ancestral topoisomerase VI. Pre-eukaryotes and pre-archaeobacteria then diverged. Once the originally thermophilic sister pre-eukaryote began to evolve phagotrophy and perfect the cytoskeleton and endomembrane system, it reverted rapidly to mesophily, since such environments provide immensely more food and the moderate temperatures would be more compatible with the relatively fluid cell surface that phagocytosis entails (Cavalier-Smith, 2002). In eukaryotes, which ancestrally had four core histones and H1, supercoiling is negative. Archaeobacteria probably never had more than two core histones (Reeve *et al.*, 1997). The archaeobacterial ancestor took the extra step into hyperthermophily, further modifying its chromatin.

Hyperthermophily and the late origins of reverse gyrase and DNA-binding protein 10b

I suggest that, as one thermophilic stem neomuran lineage adapted to hyperthermophily, thereby becoming the archaeobacterial ancestor, it lost H1 and evolved DNA-binding protein 10b, which makes negative supercoils in *Sulfolobus* only at very high temperatures (Xue *et al.*, 2000), when adapting to hyperthermophily. At the same time, reverse gyrase, found in hyperthermophiles, evolved to reduce the risk of denaturation of DNA at high temperature by supercoiling it positively. Another unrelated protein may be involved in positive supercoiling in *Sulfolobus*

(Napoli *et al.*, 2001). The argument that reverse gyrase is an evolutionary chimaera of a eubacterial DNA helicase and a eubacterial type of DNA topoisomerase I is further proof of the eubacterial ancestry and relatively later origin of archaeobacteria (Forterre, 1996). The various lines of evidence assembled here for the secondary origin of archaeobacteria leave little doubt that hyperthermophilic environments were the last major habitat colonized by free-living bacteria.

I suggest that archaeobacterial core histones, topoisomerase VI and reverse gyrase are mutually co-adapted to function efficiently. In keeping with this, *Thermoplasma*, a thermophile but not a hyperthermophile, has secondarily lost all three and replaced them, apparently by lateral gene transfer from a eubacterium, by a two-subunit eubacterial DNA gyrase (Ruepp *et al.*, 2000). However, they retained the DNA-binding protein 10b. The presence of both protein 10b and reverse gyrase in crenarchaeotes possibly allowed them to dispense with histones.

Molecular co-evolution and the origins of neomuran replication, DNA repair and transcription machinery

It has long been baffling to molecular evolutionists that the replication and transcription machinery of archaeobacteria and eukaryotes is so similar, despite their vastly different cell organization, yet so different from that in eubacteria, which have a fundamentally similar cell organization to that of archaeobacteria (despite repeated vociferous denials of this basic fact of cell biology and bacteriology by a few influential biochemists). Attributing this striking difference merely to early divergence (Woese & Fox, 1977; Woese, 1982, 2000) has always been contradicted by fossil data and reasonable interpretations of cell evolution (Cavalier-Smith, 1981, 1987b, 1990, 1991a, b); the antiquity or progenote hypothesis that maintains, contrary to such evidence, that all three domains are of equal age fails entirely to explain why neomura have one system and bacteria another. For all three reasons, the progenote hypothesis was a non-starter as a basic explanation, yet has been repeated widely, largely for want of a more convincing alternative: my arguments, that there must have been a relatively radical but rapid changeover in the transcriptional and translational machinery in stem neomurans (Cavalier-Smith, 1987b), fell largely on deaf ears. The 'dangerous question', 'why' are 'components of the Central dogma ... a package deal' (Belfort & Weiner, 1997), is best answered in terms of molecular co-adaptation between the various molecules that interact strongly with DNA and its associated proteins. I originally attributed the changeover to a genomic destabilization caused by the loss of the eubacterial murein cell wall, which is important for DNA segregation, and the sudden release of many harmful transposable elements, which I also invoked in the origin of neomuran introns (Cavalier-Smith, 1987b, 1991c, 1993).

Although such considerations might be important contributors to the origin of histones and DNA folding, they never provided a very satisfactory explanation for the radical changes in DNA replication and transcription machinery. The replacement of the eubacterial replicative polymerases by a novel type B polymerase has been particularly puzzling (Edgell & Doolittle, 1997). The simplest interpretation is that the eubacterial type B repair DNA polymerase, which can already interact with processivity factors, took over the replication function of the eubacterial PolC polymerase and underwent a gene duplication in the neomuran ancestor (Edgell *et al.*, 1998). But why should such a changeover have occurred? To get to grips with the temporally concerted changes in so many (not all) of the DNA replication and transcription enzymes, we need a fundamental evolutionary explanation. This is especially true now that it has become apparent that the same dichotomy is found for DNA repair and recombination enzymes; those of archaeobacteria are much more like those of eukaryotes than eubacteria. Thus, all four types of protein, which I shall collectively refer to as the DNA-handling machinery, underwent drastic evolutionary change at the eubacterial/neomuran transition – in whichever direction it occurred.

I now suggest that the origin of core histones in the ancestral neomurans, yielding a primitive form of chromatin, so changed the properties of the DNA perceived by many DNA-handling proteins that they also had to undergo co-adaptive changes in order to maintain high-efficiency transcription and replication. This is because DNA-strand separation is an essential part of both processes that would have been impeded and complicated by the tight wrapping of DNA around nucleosome core particles. Both initiation of nucleic acid synthesis and chain elongation would have been strongly affected. DNA repair also involves transitions between single- and double-stranded DNA or interactions between them that would have been profoundly modified by the origin of histones. The demonstration that core histones are widespread in archaeobacteria, and the deduction that they very probably evolved in the ancestral neomuran, which was not known when the neomuran theory was first presented (Cavalier-Smith, 1987b), therefore now provide a much more convincing rationale for the changeover in DNA-handling machinery than was previously possible.

Transcription. The switch from eubacterial sigma factors to the much more complex neomuran system with six interacting transcription factors was, I suggest, caused directly by the adoption of passive supercoiling of DNA by densely attached histones in place of active supercoiling by sparsely attached DNA gyrase. The tight coiling of the DNA around the histones probably necessitated a more active mechanism to initiate transcription by the TATA-box-binding protein and

associated transcription factors that mediate the binding of RNA polymerase. TATA boxes themselves probably evolved from eubacterial Pribnow boxes in the ancestral neomuran. Later, after archaeobacteria and pre-eukaryotes diverged, the RNA polymerase genes of the latter underwent triplication to make three distinct polymerases and TATA boxes were lost from the genes transcribed by RNA polymerases I and III, but retained for the majority, which use polymerase II. The multiplication of the number of RNA polymerase holoenzyme subunits from four in eubacteria to seven or more in neomura was, I argue, in turn driven by co-adaptation to the novel neomuran transcription complex and core histones. Because of the much greater complexity of the neomuran transcription machinery, it is easier to understand the origin of transcription if neomura evolved from eubacteria, as all the evidence suggests, rather than the reverse. The splitting of the second largest (A) subunit into two parts that characterizes all archaeobacteria must have occurred in a stem archaeobacterium after it diverged from the pre-eukaryote lineage; this splitting of RNA polymerase A is one of several reasons why eukaryotes are probably sisters of, rather than derived from, archaeobacteria (Cavalier-Smith, 2002). A splitting of the largest (B) subunit occurred later in the common ancestor of Neobacteria alone (Fig. 1). The presence of both splits in methanogen but not eukaryote RNA polymerases doubly refutes both recent theories of a hydrogen-using methanogen as an ancestor to eukaryotes (Martin & Müller, 1998; Moreira & López-García, 1998).

Replication. It is well known that eukaryote replication forks move about 50 times more slowly than those in eubacteria, which I have attributed to the greater difficulty of strand separation (Cavalier-Smith, 1985b). However, in archaeobacteria, the speed is more like that in eubacteria, probably because they have only two core histones and lack H1. It is sometimes suggested, because of low conservation of DNA polymerase sequences (Doolittle & Edgell, 1997), that the DNA-replication machinery evolved independently in eubacteria and neomura (Leipe *et al.*, 1999). Given the overwhelming evidence presented here for a relatively recent transition from eubacteria to neomura, this is simply not credible. The pattern of replication, bi-directional from a single origin in a circular chromosome, is identical in archaeobacteria and eubacteria (Myllykallio *et al.*, 2000). Cann *et al.* (1999) have shown that the machinery for DNA replication is fundamentally the same in all three domains. It consists of a catalytic DNA polymerase and a sliding clamp that ensures its processivity by moving along the DNA with it. In neomura, the clamp is PCNA (proliferating nuclear antigen) and its archaeobacterial homologue, a torus-shaped molecule consisting of three identical subunits. In eubacteria, the beta subunit of the replicative polymerase, which has little sequence

similarity to PCNA, forms the sliding clamp. As both clamps have an almost identical three-dimensional structure (Kong *et al.*, 1992; Krishna *et al.*, 1994), it seems virtually certain that one evolved from the other; naturally, I suggest the eubacterial version was ancestral.

It appears that the ancestral neomuran replaced the aphidicolin-resistant type C replicative DNA polymerase (pol III) alpha subunit by an aphidicolin-sensitive type B polymerase. Such aphidicolin-sensitive polymerases are not only found in several bacterial viruses, but have a scattered distribution in bacteria: in *Escherichia coli*, as a less processive repair polymerase (pol II), in some cyanobacteria and in the thermophilic posibacterium *Bacillus caldotenax* (Burrows & Goward, 1992). As they are fairly widespread as bacterial repair enzymes, there is no need to invoke a viral origin. A bacterial repair polymerase could simply have replaced the normal replicative polymerase. The repair polymerase might have proved even better than the old replicator for handling DNA wound round histones and been positively selected for that reason. I postulate that the origin of histones stimulated marked changes to the sliding clamp to allow it to continue to function properly. Such changes possibly reduced its interaction with the original pol III alpha subunit and caused it to interact more efficiently with the type B repair polymerase instead. Such direct interactions between the PCNA sliding clamp and B-type polymerases have been demonstrated (Bruck & O'Donnell, 2001). This histone-triggered co-evolution not only explains why the changeover occurred, but allows an intermediate stage in which both DNA polymerases may have been able to interact to some degree, suggesting that a smooth functional transition would have been possible without death of the cell.

The neomuran replication factor C is a heteropentameric complex responsible for loading the PCNA sliding clamp onto primed DNA. As homologues are not known in eubacteria, I suggest that this factor also evolved radically in co-adaptation to PCNA because of the origin of histones.

Repair. There are several repair enzymes unique to neomura, e.g. flap endonuclease I (FEN-1), Rad2, RadA (archaeobacteria)/Rad51 and Dmc (eukaryotes). As FEN-1 shares an octapeptide involved in binding to the interdomain region of PCNA with neomuran PolB DNA polymerases and several other eukaryotic proteins, it is clear that it is co-adapted specifically to function with PCNA. I suggest that all the novel neomuran repair enzymes and topoisomerases arose directly or indirectly as a result of the evolutionary origin of histones and the co-adaptive changes in other DNA-handling proteins such as PCNA.

There is no necessity to argue that the unique properties of the neomuran proteins reflect an early divergence from eubacteria. The weight of evidence compels us to accept that it was a secondary changeover, not a primary divergence; the fossil evidence

reviewed below suggests that the changeover was remarkably recent.

Although I argue that histones were originally an adaptation to thermophily, they can also work perfectly well in mesophiles and therefore need not be lost during secondary reversion to mesophily; in fact, they have not been lost in secondarily mesophilic euryarchaeotes (Halomebacteria and others), which, as argued above, are undoubtedly derived. The novel neomuran replicative and transcription machinery, though needed for DNA with histones, also need not undergo reversion to the eubacterial type, and would be incapable of such a reversal except by highly improbable massive simultaneous lateral gene transfer. Thus, having evolved such machinery, neomura were stuck with it, even in lineages that later lost histones; this is known to be true for the archaeobacteria [*Thermoplasma* (Ruepp *et al.*, 2000) and crenarchaeotes], but awaits explicit testing in dinoflagellates. On this molecular co-adaptive interpretation, the difference between the eubacterial and neomuran transcription and replication machinery itself polarizes the direction of evolution from eubacteria to neomura; the comparative evidence indicates that histone loss was not accompanied by a major change in this machinery, whereas histone gain was accompanied by such a change; the biophysical argument that strand separation is more difficult when histones are present explains mechanistically why this is so. It is important to stress that this co-adaptive explanation for the radical evolutionary transformation of the DNA-handling machinery is independent of the correctness or otherwise of my hypothesis that histones originated as an adaptation to thermophily. The two subtheories are logically independent and can stand or fall alone. I have presented them together because I think both are probably true; if both are, then the origin of the neomuran transcription/replication novelties was, indirectly, partially caused by ancestral neomuran thermophily.

Not every detail of this machinery need have been directly adaptive. On a reasonable view of molecular co-adaptation, it need not be. It is likely that, sometimes, one molecule becomes co-adapted to a molecular feature of another that became fixed in the first place by drift or mutation pressure; the evolution of differential intron splicing is a probable example of selection for intermolecular interactions on molecular features that originally spread by transposition pressure. The phenomenon of genetic hitch-hiking (Barton, 2000), which is easily demonstrated experimentally in bacteria and probably occurs in the wild (Tenaillon *et al.*, 1999), means that a selective sweep in response to one strong selective pressure may easily cause a neutral or even mildly deleterious mutation to become fixed indirectly; if both mutations are in the same gene (e.g. RNA polymerase), the hitch-hiking effect is particularly strong even in a population with active recombination. Compensatory base changes maintaining pairing in RNA are probably mostly examples of the

phenotypic correction of mildly deleterious mutations that may have spread initially by drift or hitch-hiking. There is no reason to think that the evolution of proteins is immune to such inevitable basic evolutionary forces. If early neomura were largely clonal, hitch-hiking could allow neutral mutations to spread; subsequent co-evolution of interacting molecules favoured by selection could stabilize originally neutral mutations that had spread on the backs of those selected directly. I have stressed this because, although I have sought to find selective explanations for the origins of all major neomuran novelties, I do not wish to argue that every molecular detail was selected directly. Many may have arisen through a complex interplay of mutational, selective, neutral, hitch-hiking and selfish principles acting on macromolecular complexes where direct physical interactions mean that they cannot evolve independently, subject to only one evolutionary force at a time, as assumed in the more simplistic models.

Co-adaptation of the DNA-handling proteins to the origin of histones in response to selection for thermophily simply solves the central conundrum of cell evolution, as Edgell & Doolittle (1997) dubbed it: the fact (surprising to them) that the largest quantum shift in DNA-handling machinery in the history of life occurred not in the ancestral eukaryote but in the ancestral neomuran. Contrary to the preconceptions of many molecular biologists, this major suite of macromolecular changes was not selected to allow the evolution of the more complex eukaryotic cell; selection has no such foresight. Instead, it was selected to make a prokaryote a more efficient thermophile. As I have long argued, the origin of eukaryote complexity was not caused by innovations in the gene-expression machinery, but by the origin of the cytoskeleton and of cytosynthesis (membrane budding and fusion; Cavalier-Smith, 1975, 1987b, 2002); an obsession with gene expression has prevented molecular biologists from understanding cell evolution, for which novel properties of gene products are fundamentally more important.

Small nucleolar (sno) RNAs – another neomuran adaptation to thermophily

Another novelty probably related to thermophily was the origin of extensive rRNA and some tRNA methylation by C/D-box snoRNAs in archaeobacteria (absent from eubacteria); such methylation appears to be more extensive in hyperthermophiles (Omer *et al.*, 2000). Whether there is also eukaryotic-like extensive pseudouridylation by H/ACA-box small nucleolar ribonucleoproteins (snoRNPs) in archaeobacteria is not known, but the evidence that their pseudouridine synthetases are more like those of eukaryotes that do use such snoRNPs (Watanabe & Gray, 2000) suggests that they may turn out to have them. If they do, such extra pseudouridylation might also initially have been an adaptation to thermophily, since pseudouridylation significantly rigidifies RNA (Charette & Gray, 2000),

which might have been particularly beneficial to hyperthermophiles. Note that these markers of pseudouridylation sites are not ribozymes, they are simply base-pairers, a very simple property of RNA. Contrary to the assumption of Poole *et al.* (1999), they could easily have evolved at any stage in evolution and need not have done so early.

Most RNA complexity and ribozymes are derived

It is often assumed and sometimes explicitly asserted (Poole *et al.*, 1999) that all ribozymal functions are relics of a hypothetical RNA world (Gilbert, 1986). This is illogical because, if RNA has an inherent ability to evolve RNA catalysis, there is no a priori reason why it could not have done so polyphyletically, in which case some examples may be phylogenetically early and others late. Arbitrarily defining the presence of genomic characters more prevalent in eukaryotes as ancestral makes it a logical necessity that eukaryotes are 'ancestral'. It is thus circular reasoning to assert that the greater prevalence of ribozymes in eukaryotes requires us to root the tree of life on them or, alternatively, on hypothetical organisms having their genomic but none of their cellular properties. The latter, purely imaginary, organisms would not be eukaryotic and it is nomenclaturally confusing and phylogenetically tendentious to call them so (Poole *et al.*, 1999). It is not an independent deduction from the facts, but a simple restatement of the phylogenetically questionable, and I argue false, assumption that all ribozymal activity must be monophyletically descended from a pre-protein world. The RNA world is a purely speculative phylogenetic hypothesis; we cannot use such an uncorroborated hypothesis, plus a logically untenable assumption, to root the tree objectively! We do not know if an RNA world ever existed (Cavalier-Smith, 2001); some chemists think it likely that RNA replaced an earlier polynucleotide (XNA) shortly after that had invented protein synthesis and protein catalysis provided the first ribonucleotides (Orgel, 1998; Nelson *et al.*, 2000). The sequence XNA world, XNA-protein world, RNA-protein world, DNA-RNA-protein world is as plausible as the RNA world hypothesis at present.

To root the tree, we must use all the molecular, cell-biological and palaeontological evidence. Poole *et al.* (1999) castigate 'fragmented approaches' and advocate 'a single continuous theory'. But their analysis is itself fragmented, as it ignores all cell-biological and palaeontological evidence and also all molecular evidence except that relating to ribozymes. Doing that is almost certain to give the wrong answer. Their advocacy of 'a single continuous theory' is a rhetorical device, making it appear that ribozyme monophyly (single continuous theory) must be true and polyphyly (fragmented approach) false. To distinguish these a priori equally reasonable possibilities, we need actual phylogenetic evidence about the origins of every kind of ribozyme to establish whether it is ancient or derived, related to others or not. The structure of the

three best-known ribozymes does not support a common origin (Herschlag, 1998). As they are associated with a virus, a viroid and a mobile type of intron, I suggest that all three evolved after the origin of protein synthesis and originated not by cellular selection but independently in different selfish genetic parasites. The tRNA-cutting function of RNase P, the only truly cellular ribozyme, suggests strongly that it evolved after protein synthesis had started and, in becoming perfected, required more precisely trimmed tRNAs than earlier; this may have facilitated the basic differentiation between chromosomes and functional RNA molecules even before cells arose (Cavalier-Smith, 1987a, 2001).

Poole *et al.* (1999) are also fragmentary in one-sidedly citing the literature on the origins of spliceosomal introns. They cite only those, like Gilbert (1986), who once believed they were all early, and none of those who have since demonstrated that they are phylogenetically late (Logsdon, 1998; Stoltzfuss *et al.*, 1997) and almost certainly evolved from group II self-splicing introns, which have now been shown to be retrotransposable mobile elements (Cousineau *et al.*, 2000), during or after the origins of mitochondria and nuclei (Cavalier-Smith, 1985c, 1991c; Roger *et al.*, 1994). This means that the RNA catalytic ancestors of the spliceosome are present in bacteria, not absent as Poole *et al.* (1999) assert; thus, on the 'eukaryotes late' view, group II introns are not a late invention but may be very ancient, dating back at least to the common ancestor of Proteobacteria and Posibacteria, which both have them (unless they underwent more recent lateral transfers). Thus, the origin of spliceosomal introns involved the recruitment of extra proteins to a pre-existing posibacterial ribozyme; if the starting assumption of Poole *et al.* (1999) was correct, this would strongly favour 'eukaryotes late', the opposite of what they assert.

Poole *et al.* (1999) ignore the selfish RNA-DNA evolutionary considerations that tell us that genomes can readily acquire vast numbers of transposable elements and therefore become much more complex than their simpler ancestors. They puzzle over why all those scores of snoRNAs should have been acquired for nucleolar processing of rRNA since bacteria get by without them. Perhaps this puzzlement can be removed on the 'eukaryotes late' view in the same simple way as it was for introns. Conceivably, they were initially selfish mobile RNA (Cavalier-Smith, 2002); but whether their spread was partially selfish or purely organismally selected, snoRNPs are fully consistent with eukaryotes late. The persistence of some introns and many snoRNPs in the cryptomonad nucleomorph (Douglas *et al.*, 2001; Maier *et al.*, 2000), which is more strongly streamlined than any bacterial genome (Zauner *et al.*, 2000; Maier *et al.*, 2000; Douglas *et al.*, 2001), refutes the assumption that both could be readily lost if bacteria had evolved from eukaryotes (Poole *et al.*, 1999), as does the discovery of an extensive methylating snoRNA system in archae-

bacteria (Omer *et al.*, 2000). Most snoRNAs are not ribozymal, but are markers of sites for methylation or pseudouridylation (Smith & Steitz, 1997); which sites they mark depends simply on base pairing, so could be readily acquired independently. Cleavage snoRNAs may be ribozymes, I suggest, but the possibility that the protein part of their RNPs is the catalyst has not been excluded. If their RNA is ribozymal, it might have evolved from RNA of RNase P, which cleaves tRNA; even if it evolved *de novo* in the first eukaryote (or an ancestral neomuran if archaeobacteria eventually prove to have them), RNA cleavage is not a novel function for a ribozyme, and this would be the slenderest of possible grounds for the cell-biologically absurd view that bacteria evolved from eukaryotes. Poole *et al.* (1999) also misconstrue the significance of eukaryotic telomerase. Contrary to their assertions, it is not a ribozyme; the catalysis is by a protein (Counter *et al.*, 1997) and the RNA is simply a template. As any RNA molecule can be a template, novel templating functions could arise polyphyletically much more easily than ribozymal functions, which do require specific sequences. Their statement that bacteria have no linear genomes is also false; some do and have clearly acquired them polyphyletically (Bendich & Drlica, 2000). Thus, all examples given of ribozymes that would be acquired secondarily on the 'eukaryotes late' view (necessitated both by cell biology and palaeontology) are false, so their 'eukaryotes early' argument falls to the ground.

Neomuran protein-spliced tRNA introns are derived

My thesis that archaeobacterial and eukaryotic tRNA introns are homologous (Cavalier-Smith, 1991c) was met with scepticism because of slight differences in splicing mechanism. However, recent discoveries have dispelled such doubts (Belfort & Weiner, 1997). In both cases, splicing requires three different enzymes: an RNA endonuclease that generates 5'-OH and 2',3'-cyclic phosphate termini (a unique intermediate for splicing), a tRNA ligase to join the 5' and 3' ends covalently, aided by ATP and GTP, and a phosphotransferase to remove the residual 2'-phosphate by transfer to NAD. In archaeobacteria, the endonuclease is a homodimer, but in yeast, it is a heterotetramer: since two of the yeast subunits are homologous to the archaeobacterial one, they must have arisen by gene duplication of the ancestral neomuran gene during eukaryote evolution. One or both of the extra non-homologous subunits in the more complex eukaryotic endonuclease may have been added to bind it to the nuclear envelope, a known feature of the enzyme (Belfort & Weiner, 1997) absent in bacteria. The cutting mechanism is fundamentally the same in both and probably involves a separate active site for each splice junction (Belfort & Weiner, 1997). In archaeobacteria, both splice junctions are probably recognized by secondary-structure bulges in otherwise paired regions; this also applies to the 3' end in fungi and animals (opisthokonts; Cavalier-Smith, 1987c), but

the 5' end, though in such a bulge, is apparently recognized by a distinct ruler mechanism. This difference may reflect the constant location of opisthokont tRNA introns, always between the first two bases 3' to the anticodon; archaeobacteria have to remove introns not only in the anticodon loop in this very position (where most reside), but also in the anticodon stem, the extra arm and even in rRNA, which sometimes has this type of intron. Archaeobacteria may therefore need a more general mechanism not dependent on substructure of the anticodon loop. That this may also be true of plants is suggested by our recent discovery in the cryptomonad nucleomorph, a relict red algal nucleus, of a tRNA intron in the D-loop, an entirely novel location (Zauner *et al.*, 2000). Its pairing potential suggests that both cuts are made in bulges, raising the possibilities that plants may have retained a more archaeobacterium-like mechanism and that the ruler mechanism of opisthokonts might not apply generally to eukaryotes.

I argue that the unique protein-spliced tRNA introns of neomura also provide decisive evidence for the derived nature of neomurans compared with eubacteria. I have argued that the extreme shortness of these introns makes the selective advantage of eliminating any individual one virtually zero, while their usual presence in the anticodon loop requires that any deletion that might achieve this be positioned with absolute precision and therefore must be of very low probability (Cavalier-Smith, 1991c). As there are several per genome, the effective elimination of all of them, though not theoretically impossible, may never have occurred during the entire history of life. Streamlining (Doolittle, 1978) and genomic reduction undoubtedly occur quite frequently in evolution, but I am deeply sceptical that they could have eliminated these tRNA introns entirely if eubacteria had evolved from neomura rather than the reverse. Such introns cannot reasonably be regarded as dating from an RNA world, since it would be a logical impossibility for the splicing mechanism, which requires three distinct proteins, to have evolved before the origin of protein synthesis. Having any intron in the anticodon loop would probably have so complicated the origin of the genetic code as to have made it practically impossible. I have therefore argued that such introns, like all other introns (Cavalier-Smith, 1978), were secondary insertions into previously unsplit genes (Cavalier-Smith 1983a, 1991c).

As their splicing involves 5'-OH and 3'-phosphate cuts in the RNA, in contrast to 3'-OH and 5'-phosphate cuts in all other introns, it has been widely assumed that protein-spliced introns originated independently. However, this by no means follows; I have stressed that a changeover of splicing mechanism could have occurred in pre-existing introns – thus, the ancestral transposable elements from which these introns evolved may have been related to the common ancestor of all other introns, even though the present splicing mechanisms are quite different and not related. The

splicing positions would have to be conserved but the splicing mechanism need not have been. I therefore argued that protein-spliced introns evolved from self-splicing introns in tRNA genes in the common ancestor of neomura (Cavalier-Smith, 1991c); there are hints that some such introns may have residual self-splicing activity (Belfort & Weiner, 1997). I argued that the selective force for a substitute splicing mechanism was energy and nutrient economy to minimize transcriptional waste, i.e. a form of genomic streamlining. I argued that it was mutationally easier to reduce the size of the introns than to eliminate them entirely; once they became as highly reduced as they now are, the selective advantage of total elimination was immensely less. This earlier analysis (Cavalier-Smith, 1991c) is strongly supported by what we have since learned about intron evolution during the most extreme cases of evolutionary size reduction of cellular genomes known to us: chlorarachnean and cryptomonad nucleomorphs.

Nucleomorphs are drastically reduced nuclei found in two independently evolved evolutionary chimaeras of two unrelated eukaryote cells. Nucleomorphs of cryptomonads evolved from a red algal nucleus (Douglas *et al.*, 1991; Cavalier-Smith *et al.*, 1996a), taken up into the common ancestor of chromalveolates (Cavalier-Smith, 1999), which was then hugely reduced in genome size, probably by at least two orders of magnitude (Cavalier-Smith & Beaton, 1999; Beaton & Cavalier-Smith, 1999). Despite immensely strong selection for genome reduction, yielding a genome with almost no non-coding DNA between genes, and an exceptionally compact genome with 44 overlapping genes, the nucleomorph genome of the cryptomonad *Guillardia theta* retains tiny protein-spliced introns in 12 different tRNA genes (Douglas *et al.*, 2001); the novel intron in the seryl-tRNA D-loop, mentioned above (Zauner *et al.*, 2000), suggests that, in neomura, insertion of these protein-spliced introns may be an ongoing process that is able to perpetuate them indefinitely in the face of the very weak selection against them.

The limited power of selection for streamlining, caused by the greater ease of intron shortening than intron deletion, is shown equally well by the persistence of short spliceosomal introns in both types of nucleomorph. The chlorarachnean nucleomorph, which evolved from a green algal nucleus, is liberally peppered with numerous tiny spliceosomal introns, shorter than in any other organism (18 ± 1 nucleotides; Gilson & McFadden, 1996) and so uniform in length as to suggest that they have evolved a novel splicing machinery using a ruler mechanism. Chlorarachneans have even smaller nucleomorph genomes than cryptomonads (smaller than any other cellular genome), indicative of the most intense streamlining of cellular genomes that ever occurred in the history of life. A novel ruler mechanism could have been selected by its facilitation of such extreme shortening. The fact that these 'bonsaied' genomes are still riddled with min-

uscule introns strongly supports my view that genomic streamlining has historically been mutationally limited and that natural selection is not an all-powerful creator. It is interesting that the cryptomonad nucleomorph has very few spliceosomal introns, which are not constant in length and are no shorter than the shortest ones known from other protists (Douglas *et al.*, 2001). The fact that they never evolved a special ability to make them shorter still may itself be an evolutionary accident, caused by the failure of the requisite mutations ever to have occurred. Having at least an order of magnitude fewer introns, the selective advantage of modifying the splicing mechanism sufficiently radically to allow such shortening was probably also very much lower in cryptomonads. However, the fact that their 13 tRNA introns are the shortest known (7, 8 and 10 nucleotides, compared with 14–106 in other organisms) testifies to the great strength and efficacy of selection for cryptomonad nucleomorph streamlining by genomic and gene shortening, so I suspect that the limitation was primarily mutational.

The above considerations mean that the tRNA introns of neomura are almost certainly a derived character for neomura and an important synapomorphy for the clade (Cavalier-Smith, 1987b). Unlike the majority of the other neomuran synapomorphies, there is no immediately obvious reason to regard splicing by proteins as an adaptation to thermophily or hot acid; its occurrence in the neomuran ancestor might therefore have been purely fortuitous. Since group I self-splicing introns occur in tRNA genes of cyanobacteria (Paquin *et al.*, 1997; Besendahl *et al.*, 2000), the probable sisters of posibacteria (see below), and occur in phage genes in posibacteria (Landthaler & Shub, 1999), they may already have been present in at least one tRNA gene of the neomuran ancestor. The substitution of a novel protein-splicing mechanism that recognized the secondary RNA structure of the splice junctions alone and no longer depended on that of the entire intron would immediately have allowed short duplications to occur in the anticodon loop without lethal consequences. Thus, many tRNAs could thereafter have rapidly acquired such introns *de novo* rather than by insertion of pre-existing introns by transposition or gene conversion, which are probably the general spreading mechanisms for self-splicing and spliceosomal introns. A very few archaeobacterial rRNA genes have probably related introns (Burggraf *et al.*, 1993); as these are usually larger and so restricted in distribution, they might be relatively recent insertions.

We should not rule out the possibility that, when it became a thermophile, a neomuran ancestor, already encumbered by one or more self-splicing introns, suffered a further lowering of growth efficiency owing to heat destabilizing the intron secondary structure needed for self-splicing. It would be interesting to investigate the temperature sensitivity of self-splicing; if it was seriously impaired at higher temperatures, this

Fig. 3. Distortion of the small-subunit rRNA tree by hyperaccelerated nucleotide substitution rates in stem neomura and stem eukaryotes. The upper figure is a schematic representation of the rRNA tree as generally observed, rooted within the eubacterial radiation as suggested by the fossil record. The eubacterial radiation is treated as an unresolved multifurcation apart from the slightly earlier divergence of Eobacteria seen on some trees and the grouping of Cyanobacteria and Posibacteria, which is seen on most taxon-rich trees and is reasonably well supported (Hugenholtz *et al.*, 1998a, b). The relative proportions of the three stems and crowns of the tree are taken from the maximum-likelihood tree of Kyrpides & Olsen (1999), ignoring the longest branches within each domain, which are caused by secondary accelerations after the primary radiation of each (see Fig. 6). In the lower figure, the long neomuran stem is moved into the actinobacterial branch, where comparative cell and molecular biology and my analysis of the neomuran transition show it belongs; it is placed at a depth in the posibacterial tree corresponding to a divergence date of 850 My, as suggested by the fossil record – the neomuran stem is lengthened so that, despite being higher in the tree, its length represents the same degree of sequence difference between neomura and actinobacteria as in the original tree. The inset shows the radical shortening of the neomuran branch that would be needed to make the tree a more accurate temporal representation of the history of life. All three stems show temporary hyperacceleration, which is greatest for the eukaryote and least for the archaeobacterial stems. By contrast, the archaeobacterial and eukaryote crowns are accelerated only slightly compared with expectations from eubacteria.

might implicate thermophily rather than selection for shorter introns as the selective force for using proteins to splice tRNA introns.

Whatever the selective force, if my thesis of replacement of splicing by RNA by protein-splicing is correct, then the view that protein enzymes should replace ribozymes and not the reverse (Poole *et al.*, 1999) would argue for the eukaryotic/neomuran state being derived and the eubacterial being ancestral – the very opposite of their hypothesis. However, for the reasons given above and elsewhere (Cavalier-Smith, 1991c, 1993), I do not regard the postulated ancestral self-splicing introns (or any other introns) as stemming from an RNA world.

Molecular co-evolution in neomuran ribosome evolution: stretching the rRNA tree

I argued previously that genetic upheaval during the origin of neomura also caused major changes in neomuran ribosomes, both rRNA and proteins, caus-

ing them to deviate markedly from clock-like behaviour during the transition and yet again during the subsequent origin of eukaryotes (Cavalier-Smith, 1987b). I suggested that the origin of general co-translational secretion of proteins entailed, at least temporarily, more rapid co-evolutionary changes in the ribosomes themselves (Cavalier-Smith, 1991a, b). At that time, I assumed that this key evolutionary step occurred in the pre-eukaryote lineage; the new evidence discussed above shows that this major change in ribosome function occurred instead in a stem neomuran, prior to the divergence of archaeobacteria and pre-eukaryotes. I suggest that the addition of the translation-arrest domain and early neomuran modifications to the docking machinery in the membrane caused numerous co-evolutionary adjustments to other features of stem neomuran ribosomes, especially in the large ribosomal subunit, from which the nascent polypeptide emerges and which interacts most strongly with the SRP and docking machinery. A second, perhaps more important, change that occurred then

was a radical modification of translation initiation, with profound repercussions, primarily on the small subunit and its 16S RNA, as I discuss in detail below. A third cause of accelerated change in rRNA may have been the adaptation to thermophily that initiated the neomuran revolution. The additional common post-translational modifications of neomuran tRNAs might also have initially been adaptations to stabilize the tRNAs at higher temperatures that then became locked in because of co-adaptive changes in both the ribosomes and the aminoacyl-tRNA synthetases.

These three relatively sudden and pervasive changes in ribosome function would have temporarily greatly accelerated evolutionary change in rRNA sequences. This explains why the distance from the central trifurcation of unrooted small-subunit rRNA trees to the apparent base of the eubacterial radiation is greater than the mean branch lengths of the major eubacterial lineages (Fig. 3). I have long considered (Cavalier-Smith, 1987a) that the eubacterial big bang radiation (Fig. 3), seen even more clearly on the best recent trees (Hugenholtz *et al.*, 1998a, b) than on early trees (Woese, 1987), corresponds to a basal radiation of photosynthetic eubacterial cells, which the fossil record indicates originated at least 3500 My ago. If this is correct, the long stem that joins the apparent base of this radiation to the trifurcation cannot possibly represent another 3500 My, but simply a short period of vastly accelerated evolution. Thus, the overall dimensions of the rRNA tree are profoundly misleading about the actual historical timing of the transition between eubacteria and neomura, as has long been evident to anyone not seduced by the false dogma of the molecular clock (Cavalier-Smith, 1980, 1981).

I also suggested that evolution of the nuclear envelope and transport of ribosome subunits to the cytoplasm, novel nucleolar biogenesis of ribosomes and novel attachments of ribosomes to the cytoskeleton all probably forced similar co-adaptive changes on ribosomes of early eukaryotes that did not occur in their sister archaeobacteria (Cavalier-Smith, 1987a, 1991a, b). Previously (Cavalier-Smith, 1981, 1987b), I argued that the origin of a 5' guanosine cap, caused by a need to prevent translation within the nucleus, was associated with marked changes in the initiation of protein synthesis, which would have temporarily greatly accelerated ribosome evolution: these factors help to explain why eukaryotic rRNAs diverged more radically from their common ancestral eubacterial rRNA than did those of archaeobacteria. The origin of the nucleolus may have been relatively less important as a cause of such divergence than previously thought, since, as just discussed, we now know that some of its features relating to snoRNA-based modification of rRNA are shared with archaeobacteria; in so far as these novelties affected the evolutionary rate of rRNA and ribosomal proteins, any shared neomuran properties would have influenced the neomuran, not the eukaryotic, stem of the tree. The origin of mito-

chondria and the consequent presence of ribosomal proteins of two phylogenetically distinct origins might have been a fifth cause of significant co-adaptive repercussions, as the paper on eukaryote origins explains in detail (Cavalier-Smith, 2002). Ribosomal changes in initiation caused by the evolution of capping were probably also significant (Cavalier-Smith, 2002), but were superimposed on more basic neomuran changes to the eubacterial initiation mechanism.

Temporary hyperacceleration of evolution and long-stem distortion of the rRNA tree

There are, therefore, strong biological reasons for expecting rRNA genes to have evolved immensely faster for a short period during the origin of neomura and eukaryotes than either before or since. From Fig. 3, it can be seen that this temporary hyperacceleration was greater in the eukaryote stem than in the neomuran stem, in keeping with the changes in the cellular fabric affecting ribosomes also being substantially greater. Fig. 3 shows that the degree of divergence among eubacteria is greater than that between plants and animals; the branch length of the latter is based on radiate animals, where rRNA evolves at much the same rate as in green plants (Cavalier-Smith *et al.*, 1996b), not on the several-fold secondary acceleration in bilaterian animals, which vitiated an early attempt to fit rRNA trees to the fossil evidence (Knoll, 1992). Previously, I pointed out the important biological distinction between permanent several-fold increases in nucleotide substitution rates that cause long branches and a temporary hyperacceleration of evolutionary rates by several orders of magnitude that is quickly reversed by a secondary slow-down and which creates immensely long, bare stems for a few lineages on some sequence trees (Cavalier-Smith *et al.*, 1996b). What, for convenience, I distinguish as the long-stem artefact and the long-branch artefact cause similar misinterpretations of the timing of evolutionary events and also errors in tree topology caused by long-branch attraction. However, the distinction between them is very important for interpreting biological history in a realistic palaeontological framework. It makes a huge difference to such correlations whether the rates along the neomuran stem were accelerated temporarily by several orders of magnitude and then sharply slowed down, as I argue, or simply accelerated several-fold as part of the same phenomenon of the accelerations within the eukaryote and archaeobacterial bush-like radiations.

To discuss these differences properly and precisely, we need to use the distinction made by the palaeontologist and cladist Jefferies (1979) between stem and crown groups. Stem groups are early relatives of a group that diverged before its cenancestor, and which became extinct, whereas the crown group comprises the cenancestor and all its descendants. Apart from recent extinctions, as of the moa, we can only get sequences

for crown groups. Thus, all extant archaeobacteria and eukaryotes are crown neomurans. However, before their cenancestor, there would have been an ancestral lineage that would have had all the shared neomuran characters just before the divergence but none of them at the point where it diverged from an actinobacterium. The neomuran novelties must have evolved at successive points along this lineage, which I shall refer to as the neomuran stem lineage or neomuran stem for short (Fig. 3). Some molecular biologists may think that I am using the term crown incorrectly, because GenBank ignorantly uses the term 'crown eukaryotes' for an arbitrary subset of eukaryotes that have short branches on rRNA trees. That misuse of the term initiated by Knoll (1992), in apparent ignorance of its proper meaning, should be discontinued. One reason why the distinction between the stem lineage and the crown group is important is that the phenotype often changes profoundly along a stem lineage. At the base of the eukaryote stem lineage, the organism was a bacterium, at its top it was a full-blooded eukaryote. If we want to map such a tree onto the fossil record, we must realize that, although we know that the cenancestor must have been a eukaryote, we do not actually know the phenotype at any intermediate point on the stem. It is an all too common mistake to assume that an entire stem lineage would have had the phenotype of its crown. I shall return to this question of the rooting of the tree and the proper interpretation of the long stems that about the central trifurcation of the unrooted rRNA tree after reviewing the fossil evidence for the recency of neomura.

The distinction between transient, short-term hyper-acceleration and sustained, long-term, mild acceleration is also important for understanding the biological reasons for accelerated evolution, which could be very different in the two cases. I argue below that the former is often associated with extremely rapid and sudden major organismal transitions that affect numerous characters (quantum evolution), as occurred during the origin of neomura and eukaryotes, whereas the latter is much more erratic.

Derived neomuran mechanisms of translation initiation

The evolution of the novel neomuran mechanism of protein synthesis initiation is another substantial 'quantum' change that can be polarized conclusively in the direction from eubacteria to neomura and was probably also a response to thermophily. In neomura, protein synthesis is initiated by methionine, whereas eubacteria and the eubacterial symbiogenetic organelles (mitochondria and chloroplasts) use *N*-formyl methionine instead. Neomura have a novel set of elongation factors (eIF-2) in addition to the universal IF-2 factors; both kinds are involved in forming the complex of the charged initiator tRNA with the mRNA and small ribosomal subunit. Neomura alone have an eIF-2A responsible for dissociating the two

ribosomes, an eIF-2B responsible for GTP recycling on eIF2 and an eIF-5A. The latter is particularly significant as the only protein in the living world with the amino acid hypusine. Since hypusine is modified from lysine by two successive enzymic steps, effected by proteins, this is compelling evidence that hypusine and the neomuran eIF-5A that depends on it are derived neomuran characters and that the simpler eubacterial system is ancestral. Just as the initiation of protein synthesis by a hypusine-dependent mechanism must be secondary, the much greater complexity of neomuran initiation points in the same direction. The eubacterial system, depending mainly on two single polypeptide factors, is far simpler and would have been much easier to evolve during the initial evolutionary origin of protein synthesis in some unknown pre-cell. The common features of translation initiation in all three domains are increasingly apparent (Condo *et al.*, 1999; Grill *et al.*, 2000; Saito & Tomita, 1999). Even its originator now argues that the earlier contention that initiation evolved independently in eubacteria and neomura is wrong (Kyrpides & Woese, 1998a, b) and that the basic features of translation initiation are universal and arose prior to the last common ancestral cell. I outline how this may have happened elsewhere (Cavalier-Smith, 2001).

Since eIF-5A is closely related in tertiary structure (Kim *et al.*, 1998) but not sequence to both IF-1 and a eubacterial cold-shock protein, it might have evolved from either. Conceivably, it was recruited to initiation by evolving from the cold-shock protein in the ancestral neomuran thermophile, because a cold-shock response then became less important than providing extra thermal stability for the initiation complex. I suggest that a variety of heat-stable proteins were recruited in this way to provide extra thermal stability to the base-pairing between Shine–Dalgarno sequences in the non-translated leader region and the small ribosomal subunit that is almost universal in eubacteria and widespread in archaeobacteria (Osada *et al.*, 1999). I suggest that, in addition to selection for stabilizing these interactions, an ancestral stem neomuran evolved, as a fail-safe device, a scanning mechanism that enabled it to initiate translation at the first methionine codon in the messenger, even in the absence of proper base-pairing of Shine–Dalgarno sequences to 16S rRNA. Recent bioinformatic analysis (Osada *et al.*, 1999) confirms earlier hints (Keeling & Doolittle, 1995) that upstream AUG codons are virtually absent in archaeobacteria, implying that such a scanning mechanism is present in all neomurans. Experimental deletion of the upstream leader confirms that archaeobacteria can initiate correctly without it and thus have an effective scanning mechanism.

The fact that a few viral messengers can similarly be recognized correctly in eubacteria implies that germs of such a mechanism exist in all cells; I suggest that they might have accompanied the evolution of the Shine–Dalgarno system and be retained universally. Possibly, thermophily caused the ancestral neomuran

to increase the efficiency of the scanning mechanism and to make it so effective that it was no longer essential to use a special *N*-formyl-methionyl-tRNA to ensure correct initiation. If accurate initiation could then occur with ordinary methionyl-tRNAs, the special initiation *N*-formyl-methionyl-tRNA and the enzymes adding and removing *N*-formyl methionine would become dispensable for the first time and were therefore inevitably lost by mutational degradation and deletion.

This thermophily scenario, therefore, simply explains the changeover in initiation mechanism during the secondary origin of neomura; it is entirely unnecessary to assume that the differences between the two groups reflect a primary divergence (Kyrpides & Woese, 1998a, b). That was a more tenable position before it was discovered that IF-2 is present universally (Keeling & Doolittle, 1995; Kyrpides & Woese, 1998b) and functions as an elongation factor even in humans (Lee *et al.*, 1999) and that IF-1 also has neomuran homologues (Kyrpides & Woese, 1998a). The addition of eIF-2 in neomura could have been to increase the stability of initiation in the absence of a specific interaction involving *N*-formyl-methionyl-tRNA. The ancestor of eIF-2 might have been a duplicate gene of the eubacterial selenocysteinyl-specific elongation factor; their sequences are homologous (Keeling *et al.*, 1998).

Derived features of neomuran selenocysteine insertion

Another intriguing feature of neomuran protein synthesis, which is more complex than in eubacteria and therefore probably derived not ancestral, is the mechanism of co-translational insertion of selenocysteine, the twenty-second amino acid (Atkins & Gesteland, 2000; Cavalier-Smith, 2001), which is encoded by UGA, which usually signifies termination. In all cells, selenocysteinyl-specific elongation factor recognizes the tRNA by binding to it. In eubacteria, this specialized elongation factor (SelB) also binds to a stem-loop immediately following the UGA, thereby placing the selenocysteine directly in the correct position for peptide bond formation by the large rRNA ribozyme. In neomura, however, messengers for selenoproteins lack this immediately downstream stem-loop and have a more distant stem-loop, usually in the 3'-untranslated tail of the molecule (one case in an upstream coding region), which is recognized not by SelB itself but by an unrelated protein (Fagegaltier *et al.*, 2000; Copeland *et al.*, 2000) that binds to these selenocysteine insertion sequences (SECIS) (Mizutani & Fujiwara, 2000). The SECIS-binding protein then binds to SelB with its already-bound selenocysteinyl-tRNA, bringing the latter to the ribosome's P-site, where peptidyl transferase adds it to the growing polypeptide chain. I suggest that this more complex two-protein system was selected in the ancestral thermophilic neomuran because the internal coding region stem-loop was prone to thermal denaturation

and its properties could not be optimized because of constraints imposed by conserved amino acid sequences in that functionally key region. By placing the stem-loop in the untranslated tail instead, its thermal stability could be freely optimized; by having a separate SECIS-binding protein, its recognition capacity for the stem-loop could also be optimized independently of the binding properties of SelB to the tRNA as long as it could bind to SelB with an appropriate geometry. Thus, the complexification of the selenyl insertion mechanism is also plausibly interpreted as a biophysical adaptation to thermophily. The two-protein system would be more difficult to evolve in the first place than the one-protein eubacterial system and there would be no advantage in changing over to the eubacterial system during secondary mesophily. Thus, selenocysteine insertion is yet another phenomenon that unambiguously polarizes the change from eubacteria to neomura and has a reasonable mechanistic and selective interpretation on the theory of secondary thermophily. The origin of the SECIS-binding protein is obscure; it has no convincing sequence relationship to eubacterial proteins – it is possible that its tertiary structure will be more informative.

Exosomes, tRNA ends and means

In eukaryotes, several exonucleases that digest RNA from the 3'-OH end are associated with RNA-binding proteins in a particle, the exosome, composed of 11–16 proteins. Indirect evidence from their presence in a single operon in an archaeobacterium suggests strongly that archaeobacteria also have exosomes (Koonin *et al.*, 2001). As they are involved in trimming rRNA precursors cut up by the snoRNA machinery, I suggest that exosomes evolved at the same time in the neomuran ancestor as part of their more complex RNA-processing machinery. Since unwanted non-specific degradation would tend to go much faster at higher temperatures, it would have been as important for a thermophile to control its RNA degradation machinery as its proteolytic machinery to prevent it from getting out of control and rapidly degrading useful RNA. Association of nucleases in a complex particle possibly enabled such control to be more effective and harmful non-specific degradation to be reduced, perhaps by reducing accessibility to the nucleases by functional native RNA molecules, analogously to the proteasome for proteins. Thus, exosomes may have started like proteasomes as an adaptation to thermophily. Although I have treated exosomes as a neomuran character, actinobacteria should also be studied, since it is possible that, like proteasomes, exosomes arose in a thermophilic actinobacterium not in the ancestral neomuran.

Harmful exonucleolytic digestion in thermophilic posibacteria may have played a role in the evolution of CCA addition to the 3'-OH end of neomuran tRNAs. Although all *Escherichia coli* genes encode the CCA, 14 of 69 *Bacillus subtilis* genes do not, and most (17 of

18 characterized) in the actinobacterium *Streptomyces* do not. This makes the total loss of encoded CCA from the 3' ends of tRNAs in the neomuran descendants of actinobacteria easier to understand than before. As addition of CCA by a specific protein terminal transferase is more complex than simply encoding it in the tRNA, I argue that an encoded CCA is the ancestral state, which probably evolved in the pre-DNA, RNA/protein world or even earlier. Later, after the evolution of DNA enabled the coding capacity of genomes to increase and to devote more genes to increased efficiency and both genetic and phenotypic repair of defects, the evolution of terminal transferases able to repair tRNAs from which one or more nucleotides were missing, whether through accidental exonucleolytic digestion or mutation, would have allowed mutations for tRNAs eliminating the encoded CCA to spread by drift, even if mildly deleterious. Because it takes longer to repair them phenotypically (Sedlmeier *et al.*, 1994), selection for rapid growth would impede the loss of terminal CCAs or favour selection for mutations that added them to the gene itself. The relatively slow growth rate of *Streptomyces* and many other Actinobacteria compared with Proteobacteria like *Escherichia coli* may predispose them to such fixation of mildly harmful tRNA mutants. The actinobacterial ancestor of neomurans had probably lost CCA from most of its tRNA genes, possibly even all, prior to the neomuran transition. If a few remained, they could also have easily similarly degenerated, since a still poorly thermally adapted ancestral neomuran is also likely to have been a relatively slow grower. Thus, the loss of encoded CCA from all tRNAs, though secondary, need not have been adaptive and could have been the easiest of all neomuran characters to evolve.

Quantum evolution of neomuran chaperonins and origin of prefoldin

I now come to the last major difference between eubacteria and neomura. This is their markedly different types of chaperonins, multi-subunit double rings that form hollow cylinders able to enclose nascent or denatured proteins and to catalyse their folding through the hydrolysis of ATP (Ranson *et al.*, 1998). Eubacterial Hsp60 chaperonins form cylinders of seven identical subunits, whereas the homologous neomuran CCT chaperonins typically have eight subunits or nine in some crenarchaeote archaeobacteria (Archibald *et al.*, 2000), which are often not identical. The increase to nine subunits in the *Sulfolobus* lineage is clearly secondary and might have been stimulated by a gene duplication to make two dissimilar subunits in the ancestral crenarchaeotes. Independent gene duplications have also occurred within the euryarchaeotes (Archibald *et al.*, 1999) and in the ancestor of eukaryotes. The eukaryote one (CCT) with eight different subunits is fundamentally more complex and thus undoubtedly derived compared with the bacterial ones; this greater complexity almost certainly arose

from its completely new functions in chaperoning tubulin and actin assembly (Archibald *et al.*, 1999; Llorca *et al.*, 1999a). However, it is clear that the neomuran chaperonins in general have a greater propensity to evolve duplicate genes than the evolutionarily more stable eubacterial chaperonins (GroEL). If eukaryotes and archaeobacteria are sisters and derived from eubacteria, as all other evidence discussed here strongly indicates, then the eightfold neomuran chaperonin must be derived from the sevenfold eubacterial one. This change was accompanied by the loss of Hsp10, also known as GroES, which is a co-chaperonin that forms a cap to the cylinder. Concomitant loss of Hsp10 is understandable if the extra domain present in the CCT group of chaperonins acts as a built-in cap (Horwich & Saibil, 1998; Llorca *et al.*, 1999b); a chaperonin can have either a built-in cap or an attachable one, not both. I suggest that a built-in cap may be preferable for a thermophile to a dissociable one that might be more prone to separate and allow a denatured protein to escape and become segregated into and digested by the proteasome instead. Thus, the evolution of thermophily and proteasomes can also make sense of an evolution of the integral neomuran chaperonin from the dissociable eubacterial one and the loss of eubacterial Hsp10. Evolution in the reverse direction would make no sense. The retention of the CCT type of chaperonins in secondary mesophiles confirms its suitability for such conditions.

Neomura uniquely have a hexameric jellyfish-shaped prefoldin (GimC), which interacts via its six coiled-coil tentacles with nascent polypeptides and channels them directly into the CCT cylinder. The fact that such proteins are protected from other interfering chaperones implies that GimC may interact directly with CCT (Leroux & Hartl, 2000) and, I suggest, possibly also with the ribosome subunit. I therefore suggest that all three macromolecular assemblies may have undergone co-evolution during the origin of neomura. Prefoldin probably generally consists of two kinds of small proteins of very different sequence but, in *Methanobacterium thermoautotrophicum*, one of these is present as two very slightly different variants. Neither shows convincing sequence similarity to any eubacterial protein, so their ancestry is uncertain.

Hyperthermophily and the loss of Hsp70 and Hsp90 chaperones by archaeobacteria

The fact that GimC can substitute experimentally for Hsp70 chaperone (Siegert *et al.*, 2000), which almost certainly evolved in the ancestral eubacterial cell, provides a rationale for the first time for the long-apparent loss of Hsp70 by many archaeobacteria. Since GimC is exceptionally thermostable, and also seems able to channel nascent proteins more directly to CCT, it would have proved far better than Hsp70 in the hyperthermophilic ancestor of archaeobacteria. The merely thermophilic neomuran ancestor, however,

probably retained Hsp70 long enough to transmit it vertically to the secondarily mesophilic ancestor of eukaryotes (Cavalier-Smith, 2002). The loss of Hsp90 from the ancestral archaeobacterium may be identically explicable. As no archaeobacteria have Hsp90, this loss took place in the ancestral archaeobacterial lineage.

The presence of Hsp70 in a minority of archaeobacteria (all euryarchaeotes) means either that it was lost polyphyletically (Gupta, 1998a) or that it was also lost in an ancestral archaeobacterium but regained secondarily by lateral gene transfer from eubacteria (Gribaldo *et al.*, 1999). The fact that it is entirely absent from the ancestral (paraphyletic) hyperthermophiles and found only in the secondarily mesophilic halomicrobacteria favours reacquisition by lateral gene transfer; the fact that eukaryotes have both prefoldin and Hsp70 implies that, in mesophiles, Hsp70 has some advantages over prefoldin for some proteins. The Hsp70 tree suggests weakly that it may have been reacquired twice from different posibacterial groups, but the longer branches of most of the archaeobacterial sequences raise the possibility that the separation into two clades is artefactual. I suggest that this long branch may be caused by the likelihood that the secondarily reacquired archaeobacterial Hsp70 may interact with many fewer different proteins than in eubacteria; since prefoldin had previously taken over all its functions, only some of them may have been returned to the incoming Hsp70. A great reduction in the number of interacting substrate proteins has similarly been cited as a possible reason for the even more marked acceleration in the evolutionary rate of all eight chaperonin genes of the cryptomonad nucleomorph (Archibald *et al.*, 2001). However, if prefoldin, which is not encoded by the nucleomorph genome (Douglas *et al.*, 2001), is also not imported into the periplastid space, its loss might specifically have decreased the stabilizing selection acting on the nucleomorph chaperonins. The presence of Hsp70 and co-chaperone Hsp40 and Hsp20 genes in *Thermoplasma* suggests that they may have advantages over prefoldin even in moderate thermophiles; their clustering in *Thermoplasma* is consistent with either lateral gene transfer or vertical descent, since they are in the same operon in eubacteria. Halobacteria and *Methanosarcina* also have both Hsp70 and Hsp40.

Although the mitochondrial Hsp70 chaperonin probably came from the proteobacterial ancestor of mitochondria (Roger, 1999), the hypothesis that the cytosolic and ER forms of Hsp70 and Hsp90 also did so (Gupta, 1998a, b) is not convincing (Cavalier-Smith, 2002).

The evidence discussed above indicates strongly that all 18 major neomuran suites of characters are derived secondarily from the often very different characters of their eubacterial ancestors. Let us now examine the fossil evidence that this remarkable evolutionary changeover occurred far more recently than most molecular biologists imagine and also later than I once thought (Cavalier-Smith, 1987b, 1990).

Fossil evidence for the immense antiquity of eubacteria, in particular negibacteria

Life appears to be virtually as old as the most ancient sedimentary rocks at the beginning of the Archaean eon, 3.8 Gy ago. However, there is no fossil evidence whatever that archaeobacteria are as old as eubacteria, for which an age of 3.7 Gy ago is suggested by carbon-isotope evidence indicative of photosynthetic carbon fixation by ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) (Strauss *et al.*, 1992), the carbon-fixation enzyme that is biased most strongly against ^{13}C and thus enriches organic carbon in ^{13}C . Although RuBisCO has unexpectedly been identified recently in archaeobacteria, no known archaeobacteria carry out photosynthesis mediated by RuBisCO, whereas this is the usual mechanism for Proteobacteria and Cyanobacteria, the major photosynthetic negibacterial phyla. RuBisCO is also widespread in non-photosynthetic posibacteria, but is not used by the only photosynthetic posibacteria, the heliobacteria, which are heterotrophs that do not fix CO_2 . Green non-sulphur bacteria, which I argue below may be the most primitive photosynthesizers, mostly use the hydroxypropionate cycle or the reductive dicarboxylic acid cycle for carbon fixation (Ugol'kova & Ivanovskii, 2000), but at least one uses RuBisCO (Ivanowsky *et al.*, 1999). Only the green sulphur bacteria (Chlorobia) uniformly lack RuBisCO and use the reductive tricarboxylic acid (TCA) cycle instead for CO_2 fixation. These facts make it highly probable that photosynthetic negibacteria using RuBisCO had already evolved prior to 3.5 Gy ago, when ^{13}C depletion levels become comparable to modern ones (Schidlowski, 2001). Fossil stromatolites go back as far. These are large layered structures, created by bacteria living in microbial mats on the floor of shallow sea or freshwater where exceptional environmental conditions prevent animals from destroying them. Today, they are produced by filamentous gliding cyanobacteria or green bacteria that migrate upwards towards the light as thin layers of sediment accumulate. Thus, green bacteria would have been able to make stromatolites if cyanobacteria had not yet evolved, or vice versa. The fact that ^{13}C depletion is less from 3.8 to 3.5 Gy ago than subsequently (Rosing, 1999) is consistent with the idea proposed below that green non-sulphur bacteria were the only photosynthetic bacteria that had yet evolved; those that do not use RuBisCO have much lower ^{13}C depletion than those that do (van der Meer *et al.*, 2001). Schidlowski (2001) has suggested that this lower depletion results from secondary enrichment by metamorphism, as some microinclusions in apatite crystals have normal levels of depletion, but this heterogeneity in depletion levels might result at least in part from the heterogeneity in carbon-fixation machinery of the green bacteria.

Early morphological fossils from nearly 3.5-Gy-old rocks of the Warrawoona Group, Western Australia, were identified as cyanobacteria (Schopf, 1992, 1993), but this is debatable; the 'cells' in the simple un-

Fig. 4. Major features of the fossil record interpreted in the light of cell and molecular biology.

branched filaments (*Primaevifilum*) are more irregular and angular than in cyanobacterial filaments and might just be mineral particles. The single coccoid cell colony found is a little more plausibly biogenic, but need not be cyanobacterial. Unnamed simple filamentous fossils, 1–2 µm in diameter and resembling modern gliding eubacteria such as green non-sulphur bacteria, *Flexibacter* or *Beggiatoa*, found in stromatolites about 3.4 Gy old from South Africa (Walsh & Lowe, 1985) and simple bacteria-like forms (Westall *et al.*, 2001) are more plausibly biogenic. A meshwork of thinner pyritic filaments in a 3.25-Gy-old deep-sea volcanic sulphide deposit might be chemotrophic bacteria (Rasmussen, 2000). The only other plausible fossil cells from the Archaean are 2.8-Gy-old filaments from Australia (Schopf & Walter, 1983); though referred to as cyanobacteria-like, they could equally well be green non-sulphur bacteria. Methyl hopanes from 2.7 Gy ago provide evidence for cyanobacteria (Brocks *et al.*, 1999) but, as several other bacterial phyla also make hopanoids, this is not compelling.

Biological sulphate reduction depletes ^{34}S compared with ^{32}S , so can be traced back to 3.47 Gy ago (Shen *et al.*, 2001). As these most ancient Warrawoona deposits are in gypsum, which is unstable above 60 °C, the sulphate reducers were almost certainly not hyperthermophiles like *Archaeoglobus*, the only known

archaeobacterial sulphate reducer, and therefore were almost certainly eubacteria. Mesophilic sulphate-reducing eubacteria are known from only two phyla, Posibacteria and Proteobacteria. Since, as argued below, Posibacteria are probably sisters of cyanobacteria, and this divergence appears to be later than the primary eubacterial divergence, it is most probable that the sulphur in these Archaean deposits was fractionated by sulphate-reducing Proteobacteria such as the numerous deltabacterial Thiobacteria or the geobacterium *Nitrospira*, which both appear early, diverging near the base of the Proteobacteria on 16S rRNA trees (Hugenholtz *et al.*, 1998a).

The boundary between the Archaean and Proterozoic eons is set arbitrarily at 2.5 Gy ago and does not correspond with marked changes in fossils; 2.45-Gy-old tubular sheaths (*Siphonophycus*) have been called cyanobacteria, but might easily be green non-sulphur bacteria. Only from 2.15 Gy ago onwards is there a more or less continuous fossil record of cells that are very probably cyanobacteria, which dominate the fossil record right up to the end of the Proterozoic (Fig. 4). The boundary between the Proterozoic and our own Phanerozoic eon is set 543 My ago by the origin of the first hard-bodied animal fossils. Basic geological processes remained the same throughout. The Proterozoic is divided into three eras: Early

(2.5–1.6 Gy ago), when morphological fossils are relatively simple and probably mostly, if not all, cyanobacteria; Middle (1.6–0.9 Gy ago), where some fossils are larger and more complex and a small minority were previously thought to be eukaryotic; and Late or Neoproterozoic (900–543 My ago), where undoubted eukaryotic fossils are frequent and an earlier supercontinent broke up. I shall argue that all Early and Middle Proterozoic morphological fossils are actually eubacteria, probably mostly cyanobacteria, and that neomura first arose in the Late Proterozoic era.

Thus, the fossil record clearly indicates that fossil eubacteria fixing CO₂ by RuBisCO existed from at least 3.5 (probably 3.8) Gy ago and that cyanobacteria have probably existed since at least 2.2 (possibly 2.7 or 2.8) Gy ago. Sulphur-isotope ratios suggest that Proteobacteria may have been present as early as 3.5 Gy ago. A date for the origin of cyanobacteria and oxygenic photosynthesis near the beginning of the Proterozoic, 2.5 Gy ago, would fit well with three other palaeontological facts (Kasting *et al.*, 1992). First, many of the largest banded iron formations, e.g. of the Hamersley Basin, were deposited about then; these could have resulted from the seasonal fluctuations in O₂ output by cyanobacteria in the early days before enough O₂ had accumulated to oxidize continuously the regular supply of reduced volcanic gases and reduced iron eroded from continents. The rise of red beds 2.0 Gy ago shows that the atmosphere had become oxidizing then; 500 My would have been ample to accumulate enough oxygen, despite the sink represented by the volcanic gases and iron. The increased spread of isotopic ratios of sulphides and sulphates 2.3–2.2 Gy ago suggests that bacterial sulphate reduction may have increased greatly then; as the degree of isotopic fractionation depends on the environmental sulphate concentration (Canfield *et al.*, 2000), the simplest cause of this would be a marked increase in sulphate produced by the oxidation of terrestrial pyrite (iron sulphide) by the O₂ produced by cyanobacteria.

The morphological fossil record for archaeobacteria is unfortunately non-existent. Before considering their meagre chemical fossil record, I shall first discuss the superabundant fossil record of unicellular eukaryotes, which provides compelling evidence that eukaryotes are more than four times younger than eubacteria.

Eukaryotes are much younger than often thought

Convincing eukaryotic fossils are found only in the last two periods of the Late Proterozoic: the Cryogenian (850–570 My ago), where all eukaryotic fossils are microscopic and probably Protozoa, and the Vendian (565–543 My ago), where macroscopic fossils, probably soft-bodied animals, also occur. However, no fossils can be certainly identified to a particular eukaryotic phylum before the Phanerozoic Eon (543

My ago to the present). Morphological fossils indicate that foraminifera (McIlroy *et al.*, 1994), radiolaria and green algae first appear in the Cambrian, just after the onset of the notorious Cambrian explosion of skeletal bilaterian animals, 543 My ago (Brasier, 2000). In the terminal Proterozoic era, the Vendian period shows megascopic soft-bodied fossils that could all be Cnidaria (known as the ‘Ediacara’ fauna, 565–543 My ago, even though they first appear in the last Varangerian phase of the preceding Cryogenian period of the Neoproterozoic era) and hexactinellid sponge spicules (545–549 My ago; Brasier *et al.*, 1997), suggesting that the first animals evolved about 570 My ago (Brasier, 2000).

These objective dates are more recent than some suggested by backward extrapolation using molecular tree dimensions and the dubious assumption of a molecular clock. Extrapolation of sequence changes backwards from palaeontologically calibrated bifurcations is dangerous, even using the most clock-like molecules, since the biases introduced by different model assumptions, some of which cannot be tested, can lead to uncertainties in dates several-fold greater than the minimum possible age. Backward extrapolation of the age of eukaryotes from sequence trees beyond the date of fossils is also hazardous, because transient large increases in evolutionary rate could greatly inflate and saturation problems shrink the estimates. Simple visual inspection of many protein trees (e.g. Baldauf *et al.*, 2000) has few pretensions, but suggests that all these taxa are roughly equally old and diverged in a single rapid radiation that, from the proportion of the stems to later branches that can be dated, may have been only about 850 My ago.

The oldest fossils that are convincingly eukaryotic are only 800 My old (Porter & Knoll, 2000). They include flask-shaped shells with apertures that might be testate amoebae and various cysts with spines or reticulate surface sculpturing that would probably have required both an endomembrane system and a cytoskeleton, the two most fundamental features of the eukaryotic cell, for their construction. A few plausibly eukaryotic fossils are also found in the slightly older Kwagunt formation of Arizona, about 850 My ago: flask-shaped fossils (*Melanocyrtium*), a cell with a possible excystment aperture and one with some spines. However, nine other fossil assemblages dated ~ 850 My ago from North America, Europe, Asia and Australia, including the especially well-studied Bitter Springs, Beck Spring and Miroedekha deposits, have no fossils that I accept as eukaryotic; they all appear to be cyanobacteria. This suggests that the Kwagunt fossils may be slightly younger than the others or else eukaryotes had not yet spread around the world. Some possibly eukaryotic fossils (*Trachyhystrichosphaera*) have been seen in the Lakhanda formation of Russia; however, though assigned an age of ~ 950 My, this is uncertain; they might not be significantly older than the better-dated, more clearly eukaryotic fossils. I

therefore take 850 My as the most probable time of origin of eukaryotes on present evidence. Earlier estimates are seriously inflated.

The large cells that appeared increasingly around 1200 My ago in the immensely stable mid-Proterozoic era (Brasier & Lindsay, 1998) could all be bacterial, and I suggest that they are. Several eubacterial groups can form giant cells: cyanobacteria (*Prochloron* is highly vacuolated; Lewin & Cheng, 1989), proteobacteria (*Epulopiscium* can form giant polyploid cells) and posibacteria (during protoplast regeneration, *Streptomyces* can form huge, 100 µm protoplasts with hyphae emerging). None of these large fossils (sphaeromorph acritarchs) has the complex surface sculpturing or spines that hint of the presence of a eukaryotic cytoskeleton in the spiny acritarchs, which became abundant after about 580 My ago. This significant expansion in the diversity of clearly protistan fossils was, I suggest in the companion paper (Cavalier-Smith, 2002), caused by the symbiotic origin of chloroplasts immediately following the end of the world's last near-global glaciation in the Varangerian, with which it coincides. Earlier large Precambrian fossils, notably the mid to late Proterozoic 'bangiophyte red algae' (Butterfield *et al.*, 1990; Butterfield, 2000) and the 2.1-Gy-old early Proterozoic *Grypania* (Han & Runnegar, 1992), are more likely cyanobacteria; an origin for red algae or any other macroscopic eukaryotic algae much earlier than animals is entirely implausible given the large number of molecular trees that indicate that they are of approximately equal age (e.g. Moreira *et al.*, 2000; Baldauf *et al.*, 2000; Cavalier-Smith, 2002). If one tries to integrate the sequence trees and the fossil record in the most parsimonious way, the most reasonable estimate for the origin of plastids is only about 570 My ago (Cavalier-Smith, 2002). Given that concatenated protein trees (Moreira *et al.*, 2000) indicate that red algae probably diverged from green plants after glaucophytes, it is highly improbable that red algae originated much before 500 My ago. Accepting the accuracy of these identifications of 2.1- or 1.2-Gy-old fossils and thereby implicitly postulating 1600 My of cryptic eukaryotic evolution in which macro-organisms never became diverse would be evolutionarily untenable. I interpret the 1.2-Gy-old *Bangiomorpha* (Butterfield, 2000), which has the best cell preservation, as a slightly more complex than usual, *Oscillatoria*-like cyanobacterium. Although it has a few features of cell arrangement that make it resemble the red alga *Bangia* still more, I think they are almost certainly convergent and well within the capacity of a filamentous cyanobacterium to evolve. The so-called different spore sizes on separate plants might have nothing to do with the analogous situation in *Bangia*. They might simply be two different cyanobacterial species. There are no features of these fossils that require them to be eukaryotic. If *Bangiomorpha* was eukaryotic, we should be totally at a loss to explain why eukaryotes failed to diversify and leave

millions of unambiguously eukaryotic fossils earlier than 850 My ago.

Molecular clock arguments cannot accurately date the origin of eukaryotes independently of the fossil record. The dates they use to 'calibrate' the later parts of the trees come only from the fossil dates, so are not independent. All inferred dates for the prior divergence of any two groups extrapolate these rates of change backwards until the lineages merge. But, if quantum evolution occurs early on near the time of divergence, as it usually does in groups where we have a good record both before and after the radiation, such backward extrapolation will systematically overestimate the age of divergence, probably by a very large amount. The fact that many such extrapolations give greater estimates of age than the direct evidence from the fossils that I have used probably means that the assumption of uniform rates is false, not that the fossil evidence is a poor indicator of the actual dates, apart from the usual stratigraphic gaps. Critical interpretation of the morphological fossil record shows a simple bipartite story: origin of bacteria before 3.5 Gy ago (Schopf, 1994) and of eukaryotes around 850 My ago – over four times younger.

The presence of fossil steranes 2700 My ago (Brocks *et al.*, 1999) certainly does not prove the existence of eukaryotes at that time. Sterols are produced by a few eubacteria in three different groups; in myxobacteria and methylotrophs (both Proteobacteria), the sterols have a narrower range of molecular complexity than that observed in the chemical fossils (Brocks *et al.*, 1999). However, mycobacteria (belonging to new class Arabobacteria of the posibacterial subphylum Actinobacteria) have recently been shown to synthesize cholesterol (Lamb *et al.*, 1998), like eukaryotes, and are therefore the best extant candidates for the ancestors of neomura. Predatory myxobacteria, which produce C₂₇ cholestenols (Kohl *et al.*, 1983), could have been far more ecologically dominant and metabolically diverse in the eons preceding the origin of phagotrophic eukaryote predators, as could the often morphologically complex and biosynthetically highly versatile actinobacteria. Methylotrophs also produce sterols (Rohmer *et al.*, 1980). Given the huge discrepancy between the early sterane age and the very late age of unambiguously eukaryotic morphological fossils, and the fact that the very group of bacteria likely to have been ancestral to eukaryotes, and another that would have been more dominant before eukaryotes, actually make sterols, I argue that steranes are totally useless as indicators of the time of origin of eukaryotes. The coincidence of the earliest ages of steranes, of possibly cyanobacterial hopanoids and of cyanobacteria-like morphological fossils may be significant.

The only relationship between eubacterial phyla resolved by the 16S rRNA tree (though with only moderate bootstrap support; Hugenholtz *et al.*, 1998a, b) is that between Cyanobacteria and Posibacteria, which I suggest are sister groups. A similar

affinity between them is seen on several protein trees, including some for photosynthetic proteins discussed below. This congruence suggests that the relationship is real. Its consistency suggests that the two groups diverged from each other distinctly later than the primary big bang radiation of eubacteria, which is the most reasonable explanation for the lack of resolution of the branching order of the other eubacterial phyla (Hugenholtz *et al.*, 1998a, b). Both rRNA and protein trees suggest that Actinobacteria and Endobacteria diverged almost as early as the big bang itself, as they often fail to group them together as a posibacterial clade. Mycobacteria making sterols might, therefore, have originated around 2.7 Gy ago, at the same time as cyanobacteria making hopanoids; as the two biosynthetic pathways share many early enzymic steps, I suggest that these common parts had already evolved in the common ancestor of Posibacteria and Cyanobacteria, which I estimate lived roughly 3 Gy ago. An origin of unicellular cyanobacteria about 2.8 Gy ago is consistent with the appearance of filamentous cyanobacterial fossils (orders Oscillatoriales and Nostocales) about 2.1 Gy ago, since these branch significantly more shallowly on the 16S rRNA tree than do unicellular cyanobacteria. The much later origin of the heterocyst-containing cyanobacteria is shown both by fossils and by rRNA trees (Turner *et al.*, 1999). Thus, the eubacterial part of the rRNA tree and the fossil record agree in these respects; these particular conclusions are therefore probably reasonably reliable.

Cell size and steranes, the sole classical palaeontological criteria for dating eukaryote origins, are thus equally useless for that purpose: prior to about 850 My ago, they are probably simply telling us about eubacterial history. Fossil cell morphology, a much more reliable palaeontological indicator, and the relative depth on protein trees of taxa that can be morphologically identified confidently in fossils (Moreira *et al.*, 2000; Baldauf *et al.*, 2000) are both consistent with a eukaryote origin about 850 My ago, just prior to the inordinately long Sturtian glaciation. This Cryogenian period was a time of global climatic disruption with several ice ages (Hoffman *et al.*, 1998; Hyde *et al.*, 2000), the Sturtian and Varangerian being near-global in extent (Kirschvink *et al.*, 2000), that might have caused great disruption to the previously prokaryotic ecosystems and numerous opportunities to colonize temporarily depleted niches. However, I do not believe that such external factors can be regarded as primary causes of eukaryotic or archaeobacterial origins; I have long considered that the origin of eukaryotes was mutationally limited. I contend that it required such a large number of exceptionally unusual mutations and drastic changes in cell structure (Cavalier-Smith, 1987b) as to make it improbable for it to happen more often than once in a few billion years, even in ideal circumstances. If one were to rerun earth history, it might never happen again. According to my present interpretation of the fossil evidence and my attempt to combine it with that from molecular

sequence trees, there was no slow-burning fuse prior to the Cambrian explosion of animals (Brasier, 2000), since animals evolved relatively soon after the origin of eukaryotes had produced a late Proterozoic explosion of protozoa (Cavalier-Smith, 2002). The Cryogenian glaciations probably simply delayed the origin of animals, which the slightly earlier origin of the eukaryotic cell made inevitable.

Though the origin of eukaryotes was probably mutationally limited, the Cryogenian glaciations might have opened a window of opportunity for the spread of the products of the only really major innovations for 3 Gy, the simultaneous origins of eukaryotes and archaeobacteria, by ending over a billion years of ecological stability (Brasier & Lindsay, 1998) for the until-then purely eubacterial world. I have argued that their common ancestor was a thermophile. The evolution of novel thermophiles might paradoxically have been stimulated by the Cryogenian snowball Earth episodes. It is possible that the only areas where significant primary production would have been possible would be around deep-sea thermal vents and high volcanoes and other geothermal hotspots on land. Might this have positively favoured the ancestral neomuran thermophile and the origin of the hyperthermophilic archaeobacteria?

Archaeobacteria are probably also very recent

Perhaps the only reliable indicators of the age of archaeobacteria are their tetraether lipids, which are unique to them and are preserved as chemical fossils. Head-to-head biphytanyl lipids diagnostic for the tetraether lipids of hyperthermophilic archaeobacteria are known only in the Phanerozoic, up to about 150 My ago (Chappe *et al.*, 1982; Summons & Hayes, 1992; Hahn & Haug, 1986); as they are less stable than steranes, this is almost certainly an underestimate of their true age. Tail-to-tail C₂₅ isoprenoid lipids are reputedly diagnostic for methanogens, but have also been found only in the Phanerozoic (Hahn & Haug, 1986). The best fossil evidence is therefore consistent with the view that archaeobacteria are sisters of eukaryotes and evolved at about the same time (Cavalier-Smith, 1987b).

The only palaeontological argument for an earlier origin of archaeobacteria is unsound. It lies in carbon isotopic anomalies about 2.8 Gy ago; organic carbon was then more depleted in ¹³C than would be expected for an ecosystem dominated by RuBisCO-based carbon fixation. Since methylotrophy (feeding on methane by eubacteria) can cause such extra depletion, it has been suggested that it might have been caused by an ecosystem dominated by methanogens making methane and methylotrophs feeding on it (Strauss *et al.*, 1992). However, this is a very indirect argument and an early origin of methanogenic archaeobacteria is not the only possible explanation for these anomalies. As Strauss *et al.* (1992) point out, an ecosystem dominated by chemoautotrophic eubacteria can cause

similar depletion. Thus, it cannot discriminate between archaeobacteria and eubacteria.

If eukaryotes arose about 850 My ago and archaeobacteria are their sisters, it is very likely that archaeobacteria are also only about that old. Since molecular trees and the secondary splitting of the RNA polymerase genes indicate that methanogens are significantly younger than the ancestral archaeobacteria, I suggest that they are only about 800 My old. The simplest explanation for the evolution of the unique pseudomurein wall of Methanobacteriales is that its substitution for the neomuran glycoprotein was an adaptation to prevent digestion of their previously glycoprotein walls by proteases in the environment. The evolution of bilateral animals with a mouth and anus about 534 My ago would have provided a novel adaptive zone for anaerobic bacteria. The animal gut probably became a major habitat for this subgroup of methanogens relatively soon after the origin of pseudomurein pre-adapted them for this niche. Several key genes for C_1 -transfer enzymes are so similar between eubacterial methylotrophs and methanogenic archaeobacteria that they probably were transferred from one to the other (Chistoserdova *et al.*, 1998). They might have gone from eubacteria to euryarchaeote archaeobacteria, not the reverse. That would fit the view that archaeobacteria are much younger than eubacteria and that eubacteria are paraphyletic and were the first cells (Cavalier-Smith, 1987a). The molecular evolution of methanogenesis genes merits closer study because of its significance for the timing of archaeobacterial origins.

In summary, the fossil evidence indicates that eubacteria are 3.5–3.8 Gy old, that photosynthetic bacteria using RuBisCO existed 3.5 Gy ago and that mesophilic sulphate-reducing eubacteria were present 3.47 Gy ago. As Proteobacteria are the only extant phylum with both phenotypes, they probably evolved close to 3.5 Gy ago. The eubacterial radiation that produced the major extant eubacterial lineages (Fig. 3) was probably therefore some time between 3.5 and 3.8 Gy ago. Although it is likely that stem eubacteria existed at least for a period before the cenancestor radiated, it is unlikely that this period was as long as 300 My or even 100 My. I therefore conservatively estimate the age of the cenancestor as 3.7 ± 0.2 Gy. Molecular evidence discussed below suggests that cyanobacteria are significantly younger than the cenancestor. A reasonable estimate for cyanobacteria consistent with fossil and molecular data would be 2.8 ± 0.4 Gy old. If eukaryotes and archaeobacteria originated 0.85 ± 0.05 Gy ago, they are both over four times younger than the cenancestor.

Extreme weakness of the assumption of archaeobacterial antiquity

Given that the fossil record so strongly indicates that neomura are over four times younger than eubacteria, why (apart from a regrettable general ignorance of

palaeontology) is the idea of the great antiquity of archaeobacteria so widespread? Three reasons have been advanced for considering them ancient, all exceedingly weak. None withstands close scrutiny.

First was the suggestion that the presumed presence of hydrogen and methane and/or CO_2 in early atmospheres implies that methanogens have an ancient type of metabolism; however, this tells us nothing whatever about when methanogens evolved, since CO_2 and H_2 have been available from volcanic outgassing throughout the history of life. From an environmental point of view, they could have arisen either very early or relatively late; the cell wall arguments, the rRNA tree and the split RNA polymerase data discussed above mean that methanogens are a derived phenotype within archaeobacteria and cannot be really ancient.

Second, the extreme divergence of archaeobacterial from eubacterial rRNA led to the assertion that they must have diverged at the beginning of life; but this assumes that rRNA is a chronometer, which it certainly is not; we know that it varies hundreds-fold in evolutionary rate in eukaryotes; as discussed above, the great separation of archaeobacteria and eubacteria on rRNA trees is almost certainly highly exaggerated by extreme quantum evolution in the neomuran stem.

Third, the apparently large but biologically relatively trivial differences in eubacterial and archaeobacterial gene expression molecules led to the suggestion that their common ancestor was a pre-cellular progenote; this never-credible hypothesis has been thoroughly demolished by genome projects that reveal (as cell-biological common sense told us all along) that archaeobacteria are just somewhat unusual bacteria [which their unwarranted and undesirable renaming as archaea (Woese *et al.*, 1990) attempted to conceal].

Despite the lack of direct evidence or any robust arguments for archaeobacteria being ancient, fashion and dogmatic tradition are so strong that some readers may be tempted to present one or more of the following counter-arguments to the thesis of neomuran recency.

At first sight, it might seem reasonable to say that the evidence against eukaryotes being present before 850 My ago is purely negative. How can you know that all the fossils before that date are actually bacteria? Might not some of them actually be eukaryotes? The answer to this is that we don't know that they are all bacteria and we certainly cannot say this merely by inspecting them. We also need to reason through the evolutionary implications of asserting that some were eukaryotic. If any fossils of 3500, 2000 or 1500 My ago were really eukaryotes, we would have to ask whether it is really evolutionarily acceptable to argue that eukaryotes existed at any of those early dates without having given rise to cells of such morphological complexity that at least some would be entirely unambiguously identified as eukaryotic. Such a view is entirely unreasonable. Such a cell would only be a eukaryote if it had an internal cytoskeleton and

endomembrane system in addition to a nucleus. I argue that, if such a cell existed as early, say, as 3000 My ago, it would within a span of as little as 10 My be bound to have evolved at least some descendants that would have such complex cell-surface sculpturing or projections that we would unambiguously agree them to be eukaryotic. I base this assertion on one important fact and two key arguments. The fact is that every phylum of protists known to us has some representatives with this degree of cell complexity, which provides historical evidence that it is not particularly difficult for any eukaryote group to evolve such complexity. My first argument is that this ability necessarily exists in any cell endowed with a cytoskeleton and endomembrane system. The functions of complex surface structures are various: mechanical support for large cells, retardation of sinking among plankton, nutrient uptake and resistance to predators. Most, probably all, of these selective advantages for complex cell-surface sculpturing would have been present for eukaryotes from the very first. Since eukaryote phylogeny makes it virtually certain that the last common ancestor of all extant eukaryotes was a phagotroph (Cavalier-Smith, 2002), it is highly probable that eukaryote origins involved the origin of phagotrophy. Given that any eukaryote has the cytological ability to make much more complex structures than any bacterium, and the selective advantages for doing so would have existed from the outset, it would be literally incredible that a group of eukaryotes could have existed for 2800, 1000 or even 400 My prior to the sudden appearance of complex eukaryotic fossils around 850 My ago without having generated numerous fossilizable descendants with clearly eukaryotic morphology.

A second counter-argument might be: maybe your reasoning is correct, but perhaps eukaryotes did evolve that early but never fossilized for some unknown reason – perhaps there was a change of environment or organismal properties around 850 My ago that suddenly enabled them to fossilize. If no sound suggestion is made as to why this should be so, we should regard it as antiscientific special pleading of the worst kind. It would not be an attempt to explain the facts rationally, but mere hand-waving, little better than the obscurantist invoking of a miracle or creation and equally difficult to disprove. It is certainly true that not all structures fossilize equally well; however, both organic-walled and siliceous or calcareous protist cells that fossilize well are known from several different eukaryote groups throughout the Phanerozoic (after 535 My ago), and fossils that are almost certainly eukaryotes but which cannot be identified to modern groups go back to 800–850 My ago. I can think of no credible reason why they should not have been preserved from much older deposits if they had existed, since organic-walled fossils are found in over 120 separate deposits all round the world over the preceding 2000 My. Darwin's postulate of a lack of fossiliferous rocks or preservability was not then an

unreasonable 'explanation' of the Cambrian explosion as a preservational artefact. Nowadays, however, after the discovery of billions of microfossils in numerous accurately dated Precambrian rocks sampled at relatively close intervals throughout the Proterozoic, it would simply reflect ignorance of the solid historical evidence for actual changes in microbial diversity preserved in the sedimentary rocks. The much-vaunted incompleteness of the fossil record, though trivially true, is not a respectable reason for ignoring the positive things it tells us about bacterial evolution in the Precambrian, where microfossils have been found in well over a thousand different deposits.

A third counter-argument would be to agree that the fossil record provides decisive evidence for the recency of eukaryotes, but to argue that, as there is no morphological fossil evidence for archaeobacteria, the divergence between the archaeobacteria and eukaryote lineages might have been much earlier. However, if one were to argue that archaeobacteria were present as early as 2800 My ago to account for the extra ^{13}C depletion, it would not be reasonable to maintain that their sister lineage was really eukaryotic (i.e. having internal cytoskeleton and endomembrane), since, on my preceding arguments, it also should have left identifiably eukaryotic fossils in the following 2 Gy period, which is contrary to the actual evidence. Thus, at best, such a purely hypothetical lineage could only have been prokaryotic, i.e. a bacterial lineage characterized by the presence of neomuran common properties (e.g. co-translational glycoprotein secretion, histones, snoRNAs, absence of murein) but the absence of uniquely derived archaeobacterial properties (prenyl ether lipids, glycoprotein flagellar shafts); if sterols, calmodulin and regulation by serine/threonine protein kinases were also vertically inherited by eukaryotes from the actinobacterial ancestor of neomura, as I argue (Cavalier-Smith, 2002), this lineage must also have possessed those three characters. It beggars belief that such a cytologically bacterial or prokaryotic lineage with this unique combination of characters could have persisted for 2000 My waiting to evolve into eukaryotes, but failed totally to give rise even to a single surviving prokaryotic lineage. My neomuran interpretation is potentially refutable by the future discovery of such novel types of bacteria; I confidently predict that no such lineage will ever be found among the numerous presently uncultivated and rather diverse bacterial lineages (Hugenholtz *et al.*, 1998a; Dojka *et al.*, 2000). The absence of any bacteria with such an unusual combination of characters is the fundamental reason why I maintain that the eukaryotic cell must have evolved relatively rapidly after the evolution of the neomuran characters shared with archaeobacteria. This means that eukaryote and archaeobacterial cells must be essentially similar in age and have evolved almost simultaneously (Cavalier-Smith, 1987b). My present re-evaluation of the fossil record, especially correcting the previous misinterpretation of the sterane chemical fossils, makes 850 My, rather than the

1500 My that I accepted previously (Cavalier-Smith, 1987b, 1990), the most reasonable current estimate for the age of this neomuran revolution. Of course, the actual date may be marginally earlier, because sampling of the fossil record is necessarily incomplete. This date is of such great importance for understanding the evolution of life that any new fossil evidence suggestive of an earlier origin needs to be evaluated very critically.

The main reason why archaeobacteria are so widely thought to be ancient, despite the compelling evidence that they are not, seems to be the cultural persistence of the original misinterpretation of similarity coefficients of rRNA catalogues as being good chronometers and therefore indicators of archaeobacterial antiquity (Woese & Fox, 1977), ignoring the cogent criticisms of Hori *et al.* (1982) and the powerful cell-biological and palaeontological evidence against it (Cavalier-Smith, 1981, 1987b, 1991a, b). Even though complete rRNA and protein sequences, which do not support the same interpretation, later became available, the very name archaeobacteria, the widespread belief in molecular clocks, the sheer complexity of the problem and the dominance of the standard misinterpretation have together ensured that relatively few scientists (e.g. Forterre & Philippe, 1999; Philippe & Forterre, 1999; Lopez *et al.*, 1999; Brinkmann & Philippe, 1999; Gupta, 1998a, b, 2000; Poole *et al.*, 1999; Glansdorff, 2000) have cared to criticize it. As argued above in relation to long-stem distortion of rRNA trees and the uncertainty of phenotypes on stems, the prevalent misinterpretation and misrooting of sequence trees runs deeper than even these critics have fully realized. To explain how sequence trees have been misinterpreted, and why they do not conflict in any way with the palaeontological evidence for the recency of neomura, I shall outline key concepts of quantum evolution and mosaic evolution that are very important for developing a more realistic framework for understanding molecular evolution than the classical neutralist oversimplifications (Kimura, 1963; King & Jukes, 1969), which have been seriously misleading in numerous ways.

Quantum evolution, molecular 'clocks' and the repeated misrooting of the universal tree

Irrespective of which of the plausibly important factors discussed above contributed most to the hyper-acceleration of rRNA evolution during the origins of neomura, eukaryotes and archaeobacteria, the key point is that these tremendous spurts in rRNA evolution that so strikingly distort the universal tree are prime examples of quantum evolution (Simpson, 1944). Quantum evolution is the generalization by Simpson (1944, 1953) that sometimes, for short historical periods, a character evolves immensely more rapidly than before or afterwards. Simpson recognized that there is no sharp distinction between quantum evolution and ordinary accelerated evolution, since its exceptional rapidity is at the extreme end of a

continuous scale. However, the revolutionary transformation of the phenotype of a lineage in a very short time, when a new body plan arises or an old one is radically transformed, is such an important feature of megaevolution that it deserves recognition by this special term. My 1987 neomuran theory of a common origin of eukaryotes and archaeobacteria from a posibacterial ancestor applied the principle of quantum evolution, by arguing that the common features of archaeobacteria and eukaryotes (e.g. in transcription and translation) that seemed to differ so greatly from those of eubacteria had evolved very suddenly in a short period in their common ancestor, before the two groups diverged from each other, and thereafter had changed relatively much less. The former widespread neglect of this theory, which has been strongly corroborated by the subsequent discoveries reviewed here, owes much to the counterintuitive nature of quantum evolution, which contradicts deeply engrained molecular clock dogma.

Similar, even more deep-seated prejudices were inherited from pre-evolutionary ideas of a static scale of being (Lovejoy, 1960) by the first professional evolutionist, Lamarck, and reinforced by Lyell's and Darwin's needs to counter catastrophism and anti-scientific creationism by overemphasizing steady, inevitable and slow rates of change. However, the fossil record proved that this was untrue for morphology. Instead, rates are highly variable, stasis is common and the origin of major groups is often marked by exceptionally rapid (quantum) evolution, typically followed by relatively sudden radiations (Simpson, 1944, 1953) and then a subsequently much slower rate of change. The idea of a molecular clock is frequently heuristically useful, although it is empirically false and theoretically unsound (Ayala, 1999) and too often grossly misleading.

I pointed out early on that rRNA cannot possibly be a molecular clock, since nuclear rRNA, plastid and mitochondrial RNA must have evolved at different rates (Cavalier-Smith, 1980). It is now abundantly clear that all three types of rRNA evolve at two or three orders of magnitude different rates in different eukaryotic lineages (Embley & Hirt, 1998; Philippe & Adoutte, 1998; Pawlowski *et al.*, 1997; Zhang *et al.*, 1999, 2000). In eukaryotes, the wealth of morphological evidence, both of extant and fossil species, establishes this incontrovertibly. Both morphological evidence and that from protein trees (none are really clock-like, but show idiosyncratic shifts in rate in different taxa) reveal that grossly unequal rates of rRNA evolution have sometimes led to radically wrong phylogenetic conclusions (Embley & Hirt, 1998; Roger, 1999). There is every reason to think that this is equally true of bacteria. Inequalities in rates of evolution of molecules or morphology of sisters, or the total loss of molecules, organelles or other characters, mean that simple comparisons of similarity, whether by phenetic distance measures or a cladistic parsimony,

can easily give historically false conclusions. The great variations in the rate of evolution of rRNA and proteins among different lineages are now widely accepted. Much less known are the peculiar effects of quantum evolution (purely temporary hyper-acceleration), which I first attempted to explain with reference to early animal evolution (Cavalier-Smith *et al.*, 1996b). They are even more important for understanding the rooting of the universal tree.

The progenote hypothesis, both in its early (Woese & Fox, 1977; Woese & Gupta, 1981) and recent (Woese, 1998, 2000) versions, assumes that quantum evolution could have occurred only during the earliest phases of evolution, before transcription, translation and replication were 'stabilized'. This central assumption, however, is fallacious. Quantum evolution can occur at any stage of evolution and is not restricted to the earliest phase of evolution. That the notion of stabilization has value is obvious, as confirmed by the immense periods of near stasis over billions of years. But, from the fossil record, we must conclude that subsequent destabilization is possible and must have occurred to enable the late origin of neomura. The neomuran theory suggests that this destabilization was a result of three distinct but causally and temporally connected major switches in adaptive zone, each involving drastic changes in cell structure: (i) the neomuran replacement of murein by glycoprotein and the related changes in protein secretion and chaperone mechanisms and the origin of histones and novel properties for DNA-handling enzymes, probably associated with secondary thermophily; (ii) the archaeobacterial lipid and flagellar shaft replacement and acquisition of reverse gyrase associated with hyperthermophily and acidophily; and (iii) the evolution of eukaryotic properties associated with the origin of phagocytosis and reversion to mesophily (Cavalier-Smith, 2002). The progenote theory is simply wrong; I stress that that it is not the rRNA tree itself, but its temporal interpretation, that is fundamentally wrong. The fundamental mistake of the progenote theory is that it ignores palaeontology and thus all objective evidence about the timing of past evolutionary events and therefore substitutes invalid assumptions about rates at different times for actual evidence. The fossil record tells us unambiguously that the quantum evolution that generated the neomura occurred nearly 3 Gy later than Woese has persistently assumed. It is in the light of this historical fact that we must interpret the dimensions of molecular trees.

I have long argued that the rRNA tree of life suffers from grossly unequal rates of change (Cavalier-Smith, 1980) and is probably misrooted (Cavalier-Smith, 1987a, 1991a, 1992a, b, 1998) and I have suspected that this is also true of the eukaryotic tree (Cavalier-Smith, 1995, 2000a), as many others have cogently argued (Embley & Hirt, 1998; Philippe & Adoutte, 1998; Stiller *et al.*, 1998; Stiller & Hall, 1999; Roger, 1999). As many have stressed, misrooting the tree is serious because it colours our whole way of thinking.

People speak of deeply diverging groups or primitive or derived characters but, if the tree is misrooted, the inferred direction of evolution in parts of it may be the opposite of the true one. Unfortunately, misrooting the bacterial part of the tree has severe repercussions on assumptions about the nature of the bacterial ancestor of eukaryotes, as well as about the relationship between eubacteria and archaeobacteria and the nature of the first cell; thus, correct views of all three problems are partly interdependent, which has made their resolution unusually difficult. Fig. 5 explains the problem with the conventional Iwabe *et al.* (1989) rooting of the tree using protein synthesis elongation factors. The key problem is that the neomuran branch and stems within each subtree, and the other subtree as a whole, are all immensely long branches compared with those in the eubacterial bush, which therefore attract each other artefactually by the classical long-branch artefact (Felsenstein, 1978), a problem raised earlier (Cavalier-Smith, 1991) and given much recent attention by Philippe & Forterre (1999) (see also Brinkmann & Philippe, 1999; Lopez *et al.*, 1999; Forterre & Philippe, 1999). No reasonable scientist familiar with the enormity of long-branch artefacts could honestly say that they are confident that the rooting shown by the elongation factor paralogue tree is correct. I am sure that it is wrong. The other classical example of the ATPase α - and β -subunits (Gogarten *et al.*, 1989) is markedly worse; in that case, the length of the neomuran stem is about seven times that of the branches in the eubacterial crown. Since the fossil record shows unambiguously that the latter represents at least 3.5 Gy, a believer in the molecular clock would have to argue that the evolutionary phase represented by the neomuran stem endured about 25 Gy, twice the age of the universe! More likely it was well under 10 My, and both ATPase genes were evolving at well over 2000 times their normal rate for a period around 850 My ago. The neomuran stem is less grossly stretched on the tree for the signal recognition protein and SRP protein 54 paralogues (Gribaldo & Camerano, 1998), being comparable in length to that of the cyanobacterial/plastid clade, suggesting that these proteins evolved over 200-fold their normal rate during the neomuran revolution when the SRP acquired its novel translation arrest domain; but this degree of stretching is probably sufficient explanation for the root being misleadingly placed there. Most duplicated paralogue trees suffer from this long-branch problem. Those few that do not, root the tree in the eubacteria instead of in the classical, but incorrect, position in the neomuran stem (Kollmann & Doolittle, 2000). Fig. 5 also illustrates the point that EF-1 α /Tu is less able to resolve the relative branching order of eukaryotes, crenarchaeotes and euryarchaeotes than is EF2/G. The branch lengths within the three domains indicate that it is more slowly evolving than EF2/G, which means that there will be fewer synapomorphies supporting each branch. This lower resolving power and questionable interpretations of indel data account for the claims for archaeobacterial paraphyly based on

Fig. 5. Misrooting of protein paralogue trees by the long-stem artefact. Least-squares distance tree of protein synthesis elongation factors redrawn from Fig. 2 of Kollmann & Doolittle (2000) so as to root it between the Cyanobacteria and Proteobacteria, as indicated by the fossil record and the indel data shown in Fig. 7. Note the exceedingly long neomuran and eukaryote stems in both subtrees, resulting from extreme quantum evolution during the origin of neomuran and eukaryote cells. If there were no long-branch artefacts of phylogenetic reconstruction, the neomuran clade of the EF-2/G subtree (thin black branches) would be expected to be positioned at the green arrowhead on the neomuran theory that it evolved from an actinobacterial posibacterium; although it actually groups with the longest eubacterial branch, the spirochaete, at position 1 instead, the difference between these positions is totally insignificant given the complete lack of resolution of the branching order among the eubacterial subtrees and the very short stems at the base of the eubacterial radiation, relative to the terminal branch lengths. In the absence of long-branch attraction, the neomuran theory would expect the EF-1 α /EF-Tu clade (orange branches) to be at the base of the eubacterial radiation, as shown by the black arrowhead; long-branch attraction between this clade and the neomuran clade of the EF-2/G subtree is so severe that it artefactually branches from the excessively long neomuran stem at position 2 instead. The inset shows how long-branch attraction by the neomuran stem will move the observed root into it (open circle) irrespective of whether the true root (closed circle) is in the eubacterial bush, as the fossil record indicates it to be (left), or among the eukaryotes, as Forterre (1995) postulated (right).

EF-1 α /Tu (Rivera & Lake, 1992; Baldauf *et al.*, 1996), which the vast majority of other data does not corroborate. Making trees based on concatenated proteins does not solve the problem of systematic bias caused by quantum evolution if most of the included proteins are similarly biased, as is likely for the ribosomal proteins that probably dominate the trees of Teichmann & Mitchison (1999).

The analyses of Philippe & Forterre (1999) and Lopez *et al.* (1999) give strong support to the thesis that long-branch attraction caused by the great length of the neomuran stem is probably responsible for misrooting the tree there. Unfortunately, however, they call this stem the eubacterial branch, thus overlooking the important distinction discussed above between quantum evolution along a stem and long-term evolution within a branching clade. They therefore suggest mistakenly that eubacteria have an elevated evolutionary rate compared with neomura, the opposite of the correct interpretation. Although they diagnosed the problem partially correctly, they have not realized

that the fossil record provides clear evidence that their favoured solution to it is wrong. Earlier, Forterre and colleagues (Forterre, 1995; Forterre *et al.*, 1993) invoked the idea of streamlining (Doolittle, 1978) to raise the possibility that eukaryote cells were ancestral and that bacteria evolved from them by simplification. Philippe & Forterre (1999) point out correctly that, if this were true and the real root lay within the eukaryote bush, the excessive length of the neomuran stem would attract the long-branch paralogue to that position of the tree and give the observed false rooting (see Fig. 5 inset). However, they overlook the fact that, if the root were actually in the eubacterial bush, as the neomuran theory argues (Cavalier-Smith, 1987a, b), the long-branch attraction would also draw the paralogue tree away from there into the long neomuran stem. The same would be true if the root were really among the archaeobacteria, as Lake (1988) suggested. Because of the extreme distortion by quantum evolution of all the molecular paralogue trees that show three clear-cut domains, mathematical reconstruction of sequence trees alone cannot determine the root unambiguously.

Fig. 6. Serious misplacement of 'deep'-branching eubacterial and eukaryotic taxa on rRNA trees caused by lineage-specific accelerated evolution. The upper figure is a schematic representation of observed rRNA trees based on the maximum-likelihood tree of Kyrpides & Olsen (1999) with the addition of a few taxa that they did not include. Various rogue taxa, notorious for being misplaced on rRNA trees and therefore omitted from Fig. 3, are shown in orange. The lower figure shows their correct evolutionary position, as determined by protein trees and cell-biological data; in this tree their branches should have been lengthened as in Fig. 3 to compensate for their being placed closer to their actual relatives, but this has not been done, for lack of space, so the branch lengths underestimate the actual degree of artefactual stretching by accelerated evolution, base-compositional biases and covarion shifts, which are together probably responsible for the misleading rRNA trees. The correct positions of the hyperthermophilic eubacteria (*Aquifex* and *Thermotoga*) are discussed in the text and those for the rogue eukaryotes by Cavalier-Smith (2002); these gross systematic errors give an almost totally wrong picture of eukaryotic evolution. Together with the quantum evolution that generates the long stems explained in Fig. 3, these artefacts mean that the rRNA tree has hugely misled most thinking about microbial evolution over the past quarter-century. The misleading attraction of the long eukaryotic branches into the eukaryotic stem is analogous to the artefactual attraction of the relatively still longer branches of protein paralogues from the correct position within eubacteria into the neomuran stem to the position P, as explained in Fig. 5; unless the true position of the rogue eukaryotic taxa is realized, as in the lower figure, their artefactual presence in the eukaryotic stem would lead to an underestimate of its degree of quantum evolution. Note that the misplacement of Archamoebae and Mycetozoa also conceals their sister relation as members of the subphylum Conosa (Cavalier-Smith, 1998).

Philippe & Forterre (1999) show that the near saturation of the molecules used for paralogue rooting introduces random noise, and is thus a problem. However, the attempt by Lopez *et al.* (1999) to extract the 'true' root by concentrating on the most conserved and least saturated residues of the elongation factors would not give the correct answer if these are severely affected by quantum evolution. The dimensions of the tree based on the conserved residues indicates that it is dominated by quantum evolution in the neomuran stem and secondarily in the eukaryotic stem for EF2.

Thus, it still shows the root in the neomuran stem; the fact that this is supported relative to that in the eukaryotic stem by relatively fewer substitutions is a natural consequence of the pruning of the more variable data and should not be interpreted as evidence for the latter alternative, as they tend to imply.

The only way of deciding objectively between the three theories is to use the information about timing that the fossil record supplies. As I have pointed out before (Cavalier-Smith, 1987b, 1990, 1991a, b, 1992a) and in

detail above, the fossil record categorically refutes the hypothesis of Forterre (1995) by showing unequivocally the immense antiquity of eubacteria and the recency of eukaryotes. In conjunction with the conclusive evidence for the sisterhood of archaeobacteria and eukaryotes, it also clearly refutes the hypothesis of Lake (1988) and also the Iwabe *et al.* (1989) rooting of the tree. Only the neomuran theory is consistent with the fossil evidence. As stressed previously (Cavalier-Smith, 1981, 1987b, 1991a, 1992b), none of the proponents of the 'eukaryotes early' hypothesis has even tried to make any remotely plausible suggestion as to how a eukaryotic cell could have lost its endomembrane system, cytoskeleton, nucleus and mitosis. Thus, these hypotheses are cell-biologically empty as well as falsified by palaeontology. By contrast, the neomuran theory has accounted for the fundamental cell-biological changes in the reverse direction in considerable detail (Cavalier-Smith, 1987b, 1991a, b, 1992c, 1993, 2002).

Apart from having being misrooted by the early protein paralogue trees, the rRNA tree is also in error in its placement of the two hyperthermophilic eubacterial groups, Thermotogales and Aquificales. rRNA puts them closer than any other eubacteria to the neomura (Fig. 6); because of this, coupled with the misrooting of the tree, both are often (mistakenly) referred to as the most deeply branching of all eubacteria. However, their cell structure contradicts this, as do many protein trees. Ultrastructurally, *Thermotoga* is a posibacterium with a single membrane; the toga of Thermotogales is not an outer membrane like that of negibacteria, but a semi-crystalline S-layer like that found in Posibacteria and Eobacteria. Both indels (Gupta, 1998a, b) and several protein trees support the inclusion of Thermotogales with Posibacteria (Cavalier-Smith, 1991a, b, 1992b, 1998). Misled by the rRNA tree, I once regrettably grouped *Aquifex* with Thermotogales (Cavalier-Smith, 1998). However, ultrastructurally, it is a typical negibacterium, like Proteobacteria, with an outer membrane with lipopolysaccharide that is only slightly unusual in chemistry (Plotz *et al.*, 2000). Two protein RNA polymerase trees suggest that *Aquifex* is related to the ϵ -proteobacteria (Klenk *et al.*, 1999), as do cytochrome *bc* trees (Schutz *et al.*, 2000). It has the insertion in alanyl-tRNA synthetase (Fig. 7), which groups it unequivocally with Proteobacteria, Sphingobacteria and Planctobacteria (Gupta *et al.*, 1999), and a long insertion in RNA polymerase β that groups it with Proteobacteria to the exclusion of Spirochaetes, Posibacteria and Cyanobacteria (Klenk *et al.*, 1999). Aquificales differ from Sphingobacteria in lipids and in sometimes having flagella. Note that the EF2-G tree (Fig. 5) puts both *Thermotoga* and *Aquifex* in their correct positions, with Posibacteria and Proteobacteria, respectively, whereas the Ef-Tu tree artefactually groups them together. EF2 is also markedly superior to the small-subunit rRNA tree in accurately reconstructing overall eukaryote phylogeny

(Moreira *et al.*, 2000). Both RNA polymerases and EF-G should be used more extensively in bacteria to test the rRNA tree. Given this congruence of evidence, I classify Aquificales within the ϵ -proteobacteria in the subphylum Thiobacteria (Table 1). The RNA polymerase tree based on the two largest subunits (Klenk *et al.*, 1999), if rooted as argued here, places neomura within the Posibacteria, but places *Thermotoga* as sister to cyanobacteria and Posibacteria/neomura; however, the branching order of these three groups has low bootstrap support, so it does not argue strongly against a posibacterial and unbacterial *Thermotoga*.

Klenk *et al.* (1999) assume that the root of the RNA polymerase β - β' tree lies in the neomuran stem; however, this stem is over twice as long as the depth of the eubacterial bush, showing extreme quantum evolution in the same way as all six molecular trees considered by Philippe & Forterre (1999), as is expected for the biological reasons discussed above. Because of this misrooting, their tentative suggestion that the eubacterial root lies within Posibacteria, specifically within mycoplasmas, is invalid. An origin of all life from obligate endoparasites of eukaryote cells would also be evolutionarily highly implausible; it seems that, as for rRNA, the evolutionary rate of mycoplasma RNA polymerase is somewhat accelerated. Green plant chloroplasts, enslaved obligate symbionts, show a more than twofold acceleration in the RNA polymerase tree compared with other plastids and cyanobacteria (Klenk *et al.*, 1999).

I suggest that the eubacterial rRNA tree has a heavy hyperthermophilic bias, perhaps caused partly by G+C-richness and partly by elevated evolutionary rates, that pushes Thermotogales and Aquificales away from their true positions towards the archaeobacteria. In the gene-content tree, *Thermotoga* is grouped weakly with the low-G+C posibacteria (where I classify it) and with Spirochaetes (Huynen *et al.*, 1999); *Aquifex* is grouped very weakly with the neomura, but its distance from the base of the Proteobacteria is actually less than that from the base of the neomura; on this tree, its closeness to the archaeobacteria may be exaggerated by likely lateral transfer of numerous archaeobacterial genes (Aravind *et al.*, 1998, 1999). *Thermodesulfobacterium*, also found in this part of the tree, is more likely to be a rapidly evolving member of the Thiobacteria than a genuinely independent lineage. I suggest that the numerous other apparently discrete lineages of uncultured thermophilic eubacteria in the same region of the rRNA tree (Hugenholtz *et al.*, 1998a) may reflect similar long-branch/base-composition artefacts and that many of them will turn out to be related to *Aquifex* or other well-known groups and not distinct and novel phyla.

Mosaic evolution during the neomuran revolution and the reconciliation of conflicting trees

Many archaeobacterial metabolic enzymes resemble those of eubacteria much more than do the DNA-

Fig. 7. A synthetic eubacterial phylogenetic tree, showing key shared innovations and losses. Several clades are very well supported by sequence trees, as well as by morphological and biochemical characters and indels that can be treated cladistically; cladistic arguments have been used to support several additional groupings poorly resolved by sequence trees. Although it is highly probable that the root of the tree lies between green sulphur bacteria (Chlorobea) and green non-sulphur bacteria (Chlorobacteria), the precise position is uncertain. If the absence of lipopolysaccharide, flagella, gas vesicles and diaminopimelic acid in the wall are all ancestral characters for Eobacteria, as argued here, then they are the most divergent living phylum and the root shown is correct. But if these characters were all secondarily lost by Eobacteria, it need not be; in that case, they might be sisters to the cyanobacterial/posibacterial lineage instead and the root would lie either just below the divergence point of spirochaetes on the present tree or between the spirochaetes and all other organisms.

handling enzymes. One reason for this is lateral gene transfer; a fair number of eubacterial enzymes have been secondarily acquired by archaeobacteria (e.g. Doolittle, 2000), especially by secondarily non-hyperthermophiles. Likewise, some archaeobacterial genes may have been transferred to eubacteria, especially to the hyperthermophiles *Thermotoga* and *Aquifex* (Aravind *et al.*, 1998, 1999) (but the extent and direction of transfer is debatable; see below). However, much of the striking contrast between the eubacteria-like and eukaryote-like archaeobacterial genes is explicable instead by the established principles of mosaic evolution, as Forterre & Philippe (1999) also argue. The term mosaic evolution was invented by the zoologist De Beer (1954) to express the fact, thoroughly established by palaeontology for morphological evolution, that different parts of an organism often evolve at radically different rates and that these rates can change suddenly and independently; thus, a dramatic change affecting only some properties makes organisms appear as a mosaic of primitive and derived characters. Only vertical descent is involved in mosaic evolution, which must not be confused with

chimaeric evolution involving lateral transfer of phylogenetically distant genes. Mosaic evolution is certainly also true of genes; some evolve much faster than others, as do different parts of the same gene – there is no universal molecular clock for any phenotypically expressed part of the genome. Furthermore, genes may shift their rate of evolution relatively suddenly, either permanently or temporarily. To many molecular biologists, mosaic evolution seems as counterintuitive as quantum evolution, but it is just as real. Ignoring it has led to countless errors of interpretation.

We should expect the genes involved in the processes listed in Table 2 to have undergone drastic change during the origin of neomura and archaebacteria, but then to have settled down once the neomuran novelties became functionally stabilized. Thus, these genes will show quantum evolutionary shifts specifically at the eubacteria–neomuran transition (the ‘neomuran revolution’) or at the base of the archaebacteria. However, more basic metabolic enzymes would not have been subject to a radical shift in function during the origins of neomura and archaebacteria. They would

therefore not undergo quantum evolution and would retain most of their ancestral characters and therefore appear much more eubacterial in character than do those included Table 2, such as the DNA-handling enzymes.

Thus, as argued earlier (Cavalier-Smith, 1981, 1987b), archaeobacteria are a mosaic of genes that underwent quantum evolution during the neomuran revolution, and therefore mostly resemble eukaryotic genes, and more conservative genes that stayed closer to those of their eubacterial ancestors and have evolved in a roughly clock-like fashion. It is vital to distinguish this mosaicism in modes of vertical evolution from chimaerism caused by lateral gene transfer or cellular mergers. To do this for any particular gene requires very thorough phylogenetic analysis. Merely labelling genes as 'archaeobacterial' or 'eubacterial' on the basis of their overall similarity conflates two fundamentally different evolutionary phenomena (mosaicism and chimaerism – to avoid nomenclatural confusion, one should not call genetic mixtures produced by lateral gene transfer mosaics, as some do). Such conflation has led to unwarranted and biologically implausible suggestions that archaeobacteria arose as evolutionary chimaeras of two unrelated cells (Koonin *et al.*, 1997).

The contrast between the conservatism of most neomuran metabolic genes and the quantum evolution of the genes listed in Table 2, and the even more radical evolution of the novel eukaryote-specific genes discussed elsewhere (Cavalier-Smith, 2002), is probably the most important example yet identified of mosaic evolution at the gene level. This contrast is phylogenetically confusing even if one makes proper trees, because those for the molecules listed in Table 2 will have their stems stretched artefactually during the period of quantum evolution, whereas most metabolic enzymes will not. This stretching is both phylogenetically beneficial and harmful. By exaggerating the differences between groups, it enables their monophyly to be determined more readily, e.g. of the three domains shown by rRNA trees and by ribosomal proteins (Brown & Doolittle, 1997). However, by creating immensely long branches, it causes great problems in rooting the tree, because long-branch ingroups will misleadingly appear as outgroups. Furthermore, the distortion of branch lengths makes such trees totally misleading for the timing of evolutionary events if interpreted within a molecular clock paradigm. Both problems are very evident in rRNA trees (Fig. 3) and in protein paralogue trees (Fig. 5).

The drastic changes in cell walls, membranes and informational molecules, and innovations in protein secretion, during the origin of archaeobacteria from a posibacterium are also prime examples of quantum evolution. The conservation of replicon structure and most metabolism are, in contrast, examples of long-term stasis. As both occurred in the same organism at one time, they exemplify mosaic evolution.

Accepting the importance of mosaic and quantum evolution makes it easy to understand why the 66 proteins studied by Brown & Doolittle (1997) gave such conflicting trees. Thirty-one of the 32 proteins that gave the traditional Iwabe *et al.*/rRNA pattern, with eukaryotes more similar to archaeobacteria, were for proteins involved in the processes listed in Table 2(a), which underwent quantum evolution in stem neomura. By contrast, 31 of the 34 proteins that did not give this pattern were metabolic enzymes, which would not be expected to have undergone quantum evolution in the neomuran stem. In the absence of quantum evolution, mildly accelerated evolution in one or other lineage would give apparently conflicting trees for these enzymes, as was observed. Such acceleration in eukaryotes alone would make the two bacterial domains appear closest (seen for 17 proteins, including all four that were not metabolic enzymes: three ribosomal proteins, which would be expected to undergo extra changes in eukaryotes, and Hsp70, which would also have undergone extra changes because of gene duplication to form the cytosolic and ER versions). Eukaryotes and eubacteria could appear closer (seven enzymes) either because of such mild acceleration in archaeobacteria alone or because the eukaryote enzyme was a mitochondrial replacement of the original (probably both happened). Equidistance between the three groups (10 enzymes) could arise if both archaeobacteria and eukaryotes underwent four-fold acceleration (assuming that the eubacterial radiation was four times earlier than the neomuran one). These differences in relative similarity observed by Brown & Doolittle (1997) were interpreted incorrectly as conflicts in the rooting of the tree. They were not that, since their trees were essentially unrooted. In most cases, they merely placed the root in the longest stem; this would give a sensible root only if rates of sequence change were uniform across the tree. This assumption is certainly false; if we accept that there is no clock, the conflicts disappear.

In a recent study of rates of protein evolution in the three domains, Kollmann & Doolittle (2000) used eight reciprocally rooted paralogue trees. Although they recognized that quantum evolution has occurred and greatly distorted the tree for the ATPase subunits, they did not realize that this is also true, albeit a little less markedly, for all the other five trees that also gave the standard Iwabe *et al.* (1989) topology, as shown in Fig. 5 for one of them; these six trees are all for neomuran quantum-evolving properties listed in Table 2(a). The only two trees for metabolic enzymes, unsurprisingly, did not give this pattern and instead intermingled the archaeobacteria and eubacteria. Kollmann & Doolittle (2000) assumed that the standard pattern reflected vertical descent and the intermingling reflected lateral transfer. An alternative explanation for this intermingling is that it arises simply because there was less rapid evolution in the neomuran stem for these two enzymes than for the other six proteins, but still a mild acceleration in rate

for neomura, erratically distributed in degree among the archaeobacterial lineages. This is more parsimonious than assuming rampant lateral gene transfers. That the basic rate of lateral gene transfer between domains is quite low is suggested by an estimate of about 1 % into *Deinococcus* (Olendzenski *et al.*, 2000). Thus, as far as timing is concerned, the intermingled trees are much closer to the truth than the canonical rRNA tree.

Feng *et al.* (1997) also found that many more trees for metabolic enzymes intermingle archaeobacteria and eubacteria than give the Iwabe *et al.* (1989) picture. The conclusion of Kollman & Doolittle (2000), that substitution rates are similar in all three domains, is vitiated by their acceptance of the standard Iwabe *et al.* (1989) rooting. If, as argued here, this is fundamentally incorrect, and the neomuran radiation is four times younger than the eubacterial one, then it follows from their calculations that the mean rate within the neomura is about four times that of eubacteria. If the neomuran stem were to represent only, say, 10 My, the quantum acceleration in the stem would range from about 500-fold for EF2/G to about 2000-fold for the ATPase. Since we know that, in two plant lineages, mitochondrial small rRNA genes have accelerated in rate quite recently by about 1000-fold (Palmer *et al.*, 2000), this is not at all implausible. I guess that such a degree of change could easily have occurred in 10 million generations, even with moderate selective forces. Since a neomuran cell could have had 1000 generations a year, this might only be 10000 years! Thus, it would easily be possible for the actual quantum evolutionary rate to have been millions of times faster than the normal rate. Misrooting the tree can thus be very serious indeed for our appreciation of past evolutionary rates, as it can conceal huge rate variations: several-fold long-term accelerations and thousandfold or even millionfold temporary ones! Given such rate variations, the extrapolation of mean rates of amino acid substitution for a great mix of enzymes calculated from the dates of fossil vertebrates only to the entire living world (Feng & Doolittle, 1997) cannot be expected to give sound conclusions.

Using lateral gene transfer to give relative times; how to disprove the neomuran theory

Most aminoacyl-tRNA synthetases can also be 'rooted' by paralogues. The variety of trees obtained is of particular interest as there is no a-priori reason why these enzymes should all co-evolve with ribosomes and thus mirror the rRNA tree; we might expect them to co-evolve with tRNA. Thus, about half the synthetases show trees like those of metabolic enzymes, in which neomura are not resolved as a clade but are mixed up among the bacteria and, in contrast with rRNA trees, there are no long, bare neomuran stems (Woese *et al.*, 2000), just as one might expect in the absence of early neomuran quantum evolution. There are, however, several trees in which the long, bare stems indicative of

such quantum evolution are found and neomuran clades are found. In some of these, eukaryotes are sisters of archaeobacteria as expected, but in others, they are embedded within the archaeobacteria, probably just badly resolved trees. The contrast between the two types of tree confirms that, for the synthetases, there was no necessary general co-adaptive reason for quantum evolution to have occurred, but shows that for some of them it did, perhaps just by chance. A similar lack of determinism for greatly elevated rates of change is seen in eukaryotic nuclear and mitochondrial rRNA and protein genes. In some cases, one can see plausible reasons (e.g. loss of cilia for tubulins), but in others, e.g. 18S rRNA of bilaterian animals or florideophyte red algae, none is apparent (Cavalier-Smith *et al.*, 1996b). This should caution one against an oversimplistic explanation of why some genes in some lineages show such change and others do not; we must sometimes accept inexplicable historical accidents and not unnaturally shoehorn every example into a monolithic explanation. The aminoacyl-tRNA synthetase trees are also complicated by lateral transfer (Wolf *et al.*, 1999), which is of two sorts. A few genes appear to be replacements of eukaryote host genes by mitochondrial ones, of which the clearest example is valyl-tRNA synthetase. Others are straightforward lateral gene transfers and have apparently occurred between all three domains in every direction. In addition, some genes show clear examples of ancient eubacterial gene duplications, sometimes prior to the cenancestor, and differential losses, and all give evidence of inconsistent branching orders probably attributable to the normal imperfections of phylogenetic reconstruction. Thus, the aminoacyl-tRNA synthetase trees exhibit to a high degree every feature known to us that can be misleading about organismal phylogeny. Yet, nonetheless, taken as a whole, they exhibit a fair degree of congruence with rRNA trees when the problems of both types of gene are sensibly allowed for. Thus, as Woese *et al.* (2000) rightly stress, both kinds of gene are probably preserving some real organismal phylogenetic signal and the difficulties of reconstructing an organismal tree are not as great as Doolittle (1999a, b, 2000) suggests.

However, Woese *et al.* (2000) interpret the contrast between the long-stem neomuran and the mixed-up trees very differently from me. They suggest that the former enzymes arose prior to the cenancestor and give the 'true' tree, while the latter arose later and moved between the domains by lateral transfer after they separated. Clearly, this implausible idea is based on the progenote model of the simultaneous ancient separation of the three domains. If neomura evolved nearly 3 Gy later than eubacteria, as the fossil record indicates, then ancestral neomura would have acquired their synthetases vertically from actinobacteria (plus, for eukaryotes, some from the pre-mitochondrial proteobacteria by lateral transfer). I have argued that all synthetases arose prior to the cenancestor (Cavalier-Smith, 2001). Some of the mixed-up trees actually

place archaeobacterial sequences near to certain sequences of posibacteria. Lateral transfer can sometimes be used to give relative dates, like stratigraphic correlation; for example, there is no way a mammalian gene could have been transferred laterally into the common ancestor of fish, but a transfer from a fish to the common ancestor of mammals would have been possible. Given recent evidence that there are no extant primitively amitochondrial eukaryotes (Cavalier-Smith, 2002), the lateral transfer of aminoacyl-tRNA synthetases and many other proteins from the α -proteobacterial symbiont proves that α -proteobacteria had evolved prior to the origin of eukaryotes. Therefore, eubacteria had diversified into phyla and classes before the origin of eukaryotes, yet another argument disproving the antiquity of the eukaryotes, and requiring that the quantum evolutionary changes that generated them came substantially after those that generated eubacteria and had nothing whatever to do with the early evolution of cells or protein synthesis. Woese *et al.* (2000) claim, without giving examples, that there are cases where an archaeobacterial synthetase gene was inserted at a phylogenetically deep position within a eubacterial group. If this were true, it would refute my suggestion that archaeobacteria are about four times younger than eubacteria.

I suggest that all seven eubacterial phyla as defined in Table 1 are at least three times older than archaeobacteria. Therefore, I predict that no lateral transfers from archaeobacteria shared by all members of any one of those seven phyla will ever be demonstrated by good phylogenetic analysis. Conversely, it is possible that a lateral transfer from a single eubacterial group may be found that can be demonstrated to be cenacestral to archaeobacteria. Such lateral transfer may be particularly helpful for getting a relative time of origin for Spirochaetes, which have no fossils and which are positioned somewhat ambiguously on present trees (see later); it appears that *Borrelia* and *Treponema* share a phenylalanyl-tRNA synthetase of archaeobacterial origin (Woese *et al.*, 2000). This suggests that these two spirochaetes diverged from each other after the date of the archaeobacterial cenacestral. This synthetase should be studied in the deeply diverging *Leptospira*; if *Leptospira* has the same apparently archaeobacterial enzyme, this would refute my suggestion that spirochaetes are much older than archaeobacteria. If, on the other hand, the *Leptospira* gene does not branch with them but with a eubacterial-type gene, this would be consistent with an earlier origin of spirochaetes. One way of proving the antiquity of archaeobacteria and disproving my ideas would be to show that all cyanobacteria, which palaeontology shows must be at least 2.5 Gy old, acquired several genes by lateral transfer from an archaeobacterium. This would be a reasonable conclusion if all cyanobacteria share several entirely unrelated genes that otherwise are known only from one archaeobacterial clade and also all these genes branch robustly within that archaeobacterial clade. If archaeobacteria really

date from near the origin of life, as the progenote theory assumes, there should have been ample opportunity for such lateral gene transfer prior to the origin of oxygenic photosynthesis.

The fact that, on the mixed-up synthetase trees, archaeobacteria generally branch within the eubacteria, not the reverse, is consistent with the evidence that they evolved later. Only one tree (that for type I lysyl-tRNA synthetase) has eubacteria branching within archaeobacteria. Significantly, that is almost the only tree of Woese *et al.* (2000) that was not rooted by an outgroup paralogue, but to accord with their arbitrary hypothesis that the archaeobacterial genes were ancestral to the eubacterial ones. I would use the fossil evidence to place the root of that tree between the two eubacterial phyla instead, which makes the archaeobacterial genes derived. On the class II lysyl-tRNA-synthetase tree, the archaeobacterial genes nest within eubacteria as sisters to Arabobacteria (Actinobacteria). I suggest that the two lysyl-tRNA synthetases arose by gene duplication prior to the cenacestral.

The neomuran theory, although clearly refutable, is consistent with all current data known to me. The progenote theory of archaeobacterial and eukaryote antiquity, however, has already been refuted decisively by the fossil evidence combined with the phylogenetic evidence that eukaryotes and archaeobacteria are sisters.

The monophyletic origin of archaeobacteria from a posibacterium

The preceding sections presented very strong evidence that archaeobacteria are not primordial bacteria but secondary adaptations to hyperthermal habitats, in which the ancestral mesophilic acyl ester membrane lipids were replaced by the more thermostable isoprenoid ether ones (Cavalier-Smith, 1987a, b), and to acid, as eubacterial flagellin was replaced by acid-stable glycoprotein. These archaeobacteria-specific changes were relatively minor compared with the concerted changes in 19 suites of characters that created their neomuran thermophilic ancestors. The sheer scale of the 27 major changes listed in Table 2 is such that one cannot possibly accept the contention of Gupta (1998a) that archaeobacteria are polyphyletic; it is incredible that all these characters could have evolved polyphyletically. The apparent intermixing of Archaeobacteria and Posibacteria on many gene trees (specifically on those for proteins not subject to the extreme quantum evolution that makes the distinctiveness of neomura so obvious on other trees) is probably misleading. It is very likely a consequence of inaccurate trees because of variable evolutionary rates and modes in both groups and often poor taxon sampling or of lateral gene transfer giving gene trees that do not mirror organismal relationships (Doolittle, 1999a, b). Excessive trust in the mythical molecular clock and overconfidence in tree reconstruction has often led to mistaken suggestions of polyphyly or lateral gene

transfer, when only normal treeing errors are involved. Archaeobacteria are undoubtedly monophyletic and probably also holophyletic.

Although one can be reasonably confident that the ancestral archaeobacterium was an acid-tolerant hyperthermophile, it is less easy to reconstruct its metabolism, because multiple losses and switches from one type to another appear to have occurred within the phylum, because we do not know the phenotype of uncultured, highly divergent lineages and because we do not know the archaeobacterial branching order with confidence. The archaeobacterial rRNA tree (Barns *et al.*, 1996) must be viewed somewhat sceptically in view of its very unequal branch lengths and short internal stems and how misleading it has sometimes been in eukaryotes; for euryarchaeotes, the rRNA and RNA polymerase trees (Klenk & Zillig, 1994) are contradictory. Taxon-rich concatenated protein trees are badly needed for testing bacterial phylogeny. Perhaps the simplest view is that the archaeobacterial ancestor was a facultatively anaerobic heterotroph with a complete TCA cycle and respiratory system able to use sulphur, nitrate and oxygen as terminal electron acceptors (Fig. 1). The ancestor was probably a facultative aerobe able to switch between anaerobic (sulphur- or nitrate-based) and aerobic respiration. It seems clear that methanogenesis evolved later within the euryarchaeotes and that halophily is also derived. As most actinobacteria are aerobic, especially in the class *Arabobacteria*, most likely to have been ancestral to *neomura*, it is likely that archaeobacteria inherited their terminal oxidases for both aerobic and nitrate respiration (Castresana & Moreira, 1999) directly from the ancestral *neomuran* and that these were lost secondarily by obligately anaerobic lineages. On this view, the few fermentative euryarchaeotes are derived. However, sulphur reduction is unknown in Actinobacteria, which makes it possible that lateral gene transfer from a sulphate-reducing proteobacterium provided the necessary polysulphide reductase and adenylylsulphate reductase to the ancestral archaeobacteria and thus played a key role in their original adaptation to solfataric habitats. Doolittle (1998) suggested that *Archaeoglobus* got its ability to reduce sulphate by similar lateral gene transfer. However, since the euryarchaeote *Archaeoglobus* and the crenarchaeote *Pyrobaculum* can both reduce sulphate and sulphite, careful phylogenetic study is needed to verify this and to determine whether any gene transfer took place independently or in their common ancestor.

By accepting the reality of quantum evolution during the origin of *neomura* and the eukaryotic kingdoms, the fossil record and molecular trees can be reconciled very simply. Given present evidence from many sources, molecular, cellular and palaeontological, the preceding evolutionary, palaeontological and cell-biological arguments clearly refute the hypotheses of archaeobacterial antiquity (Woese & Fox, 1977; Lake, 1988; Woese, 1998) and eukaryote antiquity (Woese & Fox, 1977; Poole *et al.*, 1999; Forterre, 1995; Philippe

& Forterre, 1999) and compellingly support a relatively recent origin of *neomura* from the several-fold-older eubacteria. How then did this drastic upheaval happen?

Origin of the archaeobacterial exoskeleton and membrane lipids

According to the original *neomuran* theory, the ancestor of *neomura* was a *posibacterium* with a very thick wall that secreted many external digestive enzymes, like many bacilli. I postulated that, analogously to mycoplasmas and bacterial L-forms, it lost its peptidoglycan and, after a brief, traumatic, naked phase subject to exceptionally rapid molecular evolution, it evolved a novel glycoprotein surface coat to stabilize its cell surface, initially less rigid than the eubacterial peptidoglycan (Cavalier-Smith, 1987b). The stem *neomuran* also began to use isoprenoid lipids to rigidify its surface membrane; then, after diverging from stem eukaryotes (as then hypothesized, evolving sterols for the first time), the stem archaeobacterium fully replaced acyl ester lipids with isoprenoid ethers and rigidified the *neomuran* glycoprotein layer into a true wall as it colonized hot, acid environments closed to eubacteria. Several developments now allow the origin of archaeobacteria to be more gradual and less traumatic.

The actinobacterial ancestor of *neomura* probably had a thick peptidoglycan wall (Cavalier-Smith 1987b), not a thin one as in *Thermotoga* which, misled by rRNA trees, I temporarily considered as a possible *neomuran* sister (Cavalier-Smith, 1991a). However, the discovery that mycobacteria make cholesterol (Lamb *et al.*, 1998) makes them a much better ancestral phenotype for eukaryotes than is *Thermotoga*. The other 13 characters listed in Table 3 also point very strongly to an actinobacterial origin for *neomura*, so it is very unlikely that *Thermotoga* is as closely related to *neomura* as rRNA trees suggest. Although the role of phagotrophy in the origin of eukaryotes demands a flexible cell surface and, therefore, the loss of murein, it does not require a strictly naked ancestral phase (see Cavalier-Smith, 2002). Thus, given the similarities between archaeobacterial and *posibacterial* S-layers of paracrystalline globular proteins (Sara & Sleytr, 2000), it is likely that the origin of *neomura* involved the loss of murein and lipoprotein, but not the S-layer. I suggest that the eubacterial S-layer was converted instead into the archaeobacterial glycoprotein wall. As the eubacterial S-layer proteins are secretory proteins with a cleaved signal sequence, the change could have come about, following the loss of murein, simply by a mutation that prevented this cleavage, leaving the signal peptide to anchor the glycoprotein to the membrane to form the new wall. This means that the archaeobacterial wall originated, and the major innovation of co-translational N-linked glycosylation of wall proteins occurred, not via a fragile, naked intermediate, but as a changeover in wall structure.

The key role of GlcNAc in both neomuran glycoprotein and eubacterial peptidoglycan suggests an evolutionary link between their biosynthetic pathways (Cavalier-Smith, 1987b).

The discovery that mycobacteria make cholesterol (Lamb *et al.*, 1998) makes the origin of the eukaryotic endomembrane system more gradual and easier to understand than before (Cavalier-Smith, 2002). The actinobacterial class Arabobacteria, in which I place the mycobacteria (Table 1), is the most likely ancestral group for neomura. It therefore deserves broad and deep study in order to pinpoint from which subgroup neomura evolved and to tell us more about the origins of neomura and archaeobacteria. During its adaptation to hyperthermophily, the archaeobacterial ancestor replaced its acyl ester glycerophospholipids by more heat- and acid-stable isoprenoid ether phospholipids. The initial phase of isoprenoid biosynthesis, of isopentenyl phosphate via the mevalonate pathway, is shared by sterols and the archaeobacterial lipids. However, the evolution of the strongly rigidifying prenyl ethers enabled archaeobacteria to dispense with sterols, whereas they were retained as membrane rigidifiers by their eukaryote sisters, which also needed to retain the flexible eubacterial phospholipids to allow the membrane budding and fusion associated with phagotrophy. Thus, the divergence in membrane chemistry between the two neomuran sister groups (Fig. 2) is biologically explicable through secondary adaptation of their thermophilic common ancestor to hyperthermophily by the archaeobacteria and to mesophilic phagotrophy by eukaryotes.

Another argument for a more gradual transition is the discovery that archaeobacteria are very normal bacteria in chromosome and operon organization. Euryarchaeotes also retain the eubacterial FtsZ-based division mechanism (Møller-Jensen *et al.*, 2000). Thus, the transitional phase, in which the common neomuran characters were acquired, was not as traumatic for basic bacterial cell organization as I thought previously. The original neomuran theory was influenced strongly by early expectations that archaeobacteria might turn out to be radically different from eubacteria in most respects. The fact that their basic cell biology and metabolism are fundamentally the same as those of eubacteria proves that there was no cell-wide trauma in their history, as I had postulated. Positive selection for hyperthermophily, rather than mutation pressure (Cavalier-Smith, 1987b) or selection for antibiotic resistance (Gupta, 1998a), was probably the major force behind the quantum-evolutionary innovations in the stem archaeobacterium.

Once the neomuran ancestor had evolved, very few further innovations were needed to evolve the first archaeobacterium, the key ones being lipid replacement and the modification of flagella and tRNAs (Table 2b). However, a huge number of gene losses occurred during their adaptation to hyperthermophily. These include not only the dozens listed by Gupta (1998a),

but also others discussed above, such as loss of histone H1. As their arabobacterial ancestors were much more complex metabolically than hyperthermophilic archaeobacteria and had much larger genomes, to judge from *Mycobacterium*, it is likely that many hundreds of genes – perhaps a thousand or so – were lost (many others were lost in the ancestral neomuran, e.g. ones involved in wall synthesis or protein secretion). These would have included many involved in the synthesis of sterols, acyl esters and other complex lipids. If eukaryotes inherited their capacity to differentiate into resting cysts and spores from actinobacteria, as is likely (Cavalier-Smith, 2002), this ability, perhaps involving hundreds of genes, must have been retained in the cenancestral neomuran but lost by the ancestor of archaeobacteria, none of which are known to differentiate into spores. Thus, the origin of archaeobacteria involved first a shared neomuran thermophilic genome reduction, then a further archaeobacterium-specific hyperthermophilic reduction. Later, the secondarily mesophilic halophiles re-expanded their genome, in small part by lateral transfer from eubacteria.

In summary, a secondary hyperthermophilic origin of archaeobacteria from a heterotrophic posibacterium via an intermediate thermophilic eubacterium gives a unified evolutionary explanation for all major differences between archaeobacteria and eubacteria, but a transition in the reverse direction would be incomprehensible. Thus, the universal tree must be rooted within the eubacteria, not between them and archaeobacteria or within the archaeobacteria. This argument illustrates the great power of transition analysis in polarizing evolutionary change.

Polarizing change and rooting trees: the primacy of transition analysis and fossils

Transition analysis is the name I gave (Cavalier-Smith, 1991a) to the conceptual construction of a rational sequence of steps in converting a deduced ancestor into a different descendant and the critical analysis of their mechanistic and developmental soundness and selective advantages. In transition analysis, it is essential to make both the specific mutational steps and the epigenetic basis for the relevant morphological and molecular changes as explicit as possible; thought should be given to how any changes might affect other organismal features and overall viability. Surprisingly, many of the symbiogenetic suggestions made about the origin of cilia or nuclei, for example, appear to invoke hypothetical intermediates that would be lethal. In addition to avoiding lethal intermediates, it is desirable to specify the selective forces responsible for the spread of each intermediate stage but, in practice, it is reasonable to concentrate on doing this for the key steps. The condemnation of such an approach as mere speculation or scenario-building by many cladists is antiscientific and philosophically naïve. The emphasis on rooting and polarizing change by reference to

outgroups by cladists is perfectly logical and proper. However, transition analysis plays a key role, often an essential one, in establishing what actually is an outgroup. The facts (theories for the really pedantic) that green algae are an outgroup to land plants and that Cnidaria are to bilateral animals were established by past generations of comparative biologists doing transition analysis, often quite explicitly, sometimes implicitly, so they can now be taken as given. In the areas of the tree where outgroups are uncertain, for instance among bacteria, transition analysis is the primary way, often the only sound one, of establishing the direction of evolution, as I have attempted to show in this and earlier papers.

To demand that tree-building should come first and be firmly settled before we begin the job of transition analysis is fundamentally wrong. Progress will be faster if we alternate regularly between the two modes and apply a critical but constructive approach to each. Given also the tremendous biases and pitfalls in molecular trees, it is an illusion to think that they can show the position of the root of the universal tree reliably, even with the help of a cladistic approach (Forterre & Philippe, 1999), without the help of transition analysis of both sequence and cell-biological characters.

Fossils, of course, provide the only fairly direct evidence about actual past organisms, environments and timing of events. However, they are difficult to interpret, both because of their fragmentary nature and because they do not provide a direct picture of past phylogeny. Even if the record were perfect, it could only be converted into a phylogeny with the help of both cladistic and transition analysis. The radiometric clocks used to date fossil events can be remarkably accurate, but they give us dates only above and below fossiliferous strata; the problem of worldwide stratigraphic correlation is also by no means trivial, while the difficulty of identifying many microbial fossils is immense. Therefore, not all dates assigned to taxa in the literature can be trusted. Too often, the incompleteness of the record is used by molecular researchers as an excuse for ignoring it. But one cannot get dates from molecular trees (unless they are palm trees); they all ultimately come from palaeontology (Lee, 1999). Like Lee, I stress the importance of palaeontology, since it provides the only objective data on the timing of evolutionary events, making it an indispensable corrective to the subjective speculations and unreasonable inferences so widespread in molecular biology.

The idea that there can be a simple, objective algorithm for constructing phylogeny that necessarily gives us the truth is nonsense, especially if it relies on a single line of evidence, whether molecular or morphological. There is no substitute for thinking and weighing and evaluating often conflicting evidence. Even if we do this, we shall make mistakes, as we all do. But, with increased knowledge and careful criticism, these will be corrected, though often such corrections are re-

tarded by our human propensity to follow fashions and repeat dogmas with insufficient consideration of alternatives.

Quantum evolution and mosaic evolution in relation to the three domains of life

On very rare occasions, symbiogenesis has radically increased cell complexity, most strikingly in the origins of eukaryote algae (Cavalier-Smith, 1995, 2000a). However, it is the exception, not the rule. Neither lateral gene transfer nor symbiogenesis can explain real innovation; they can only move existing things from one place to another. Symbiogenesis played no part in bacterial evolution. Most of the increases in complexity and origins of major groups, such as archaeobacteria, spirochaetes or cyanobacteria, have involved quantum and mosaic evolution, but not symbiogenesis or lateral gene transfer. The assertion that vertical inheritance is never innovative, but lateral transfer is (Woese, 2000), is the exact opposite of the truth and all we know about the origins, for example, of eukaryotic phyla and classes; however, it helps us understand why Woese clings so firmly to his early idea that no transition by normal vertical evolution is possible between the three domains (Woese, 1982), despite the overwhelming evidence that just such a transition did occur around 850 My ago. The origin of eukaryotes was unusual in involving both vertical quantum change and lateral transfer by symbiogenesis, but even here, autogenous quantum changes caused the most radical and most numerous biologically significant innovations – the symbiogenetic origin of mitochondria was important, but much less innovative (Cavalier-Smith, 2002). In cases like this, it is incumbent on us to identify the selective forces (or, especially for genomic properties, mutational forces; Cavalier-Smith, 1991c, 1993) that caused some genes and characters to change unprecedentedly fast and others to languish in the doldrums.

Take just one case: the origin of the three tubulins from FtsZ. As Doolittle (1995) remarked, change in this molecule during the transition from bacteria to eukaryotes must temporarily have been 10–100 times faster than within bacteria or eukaryotes. Similar considerations apply to many hundreds of molecules that underwent radical innovations during the neomuran, eukaryotic and archaeobacterial transitions between the three domains of life, often so great as to obscure or even overwrite sequence evidence of their ancestry. Such a use of the term ‘domain’ is convenient and acceptable, so long as we avoid the serious mistakes of calling all three domains primary (Woese & Fox, 1977; Woese & Gupta, 1981; Pace *et al.*, 1986; Pace, 1991), unwisely denying the possibility of transitions between them (Woese, 1982) or denying (Woese, 1994, 1998) the reality of the more extensive and far more important distinction between the empires (or superkingdoms, if you prefer) Prokaryota and Eukaryota (Mayr, 1998). Nor should we refer to the domains as kingdoms, which none is in a sensible

taxonomy (Cavalier-Smith, 1998). According to the neomuran theory of the evolution of the three domains, updated here, eubacteria are the only primary (basal or paraphyletic) domain of life; archaeobacteria and eukaryotes are both secondary (terminal or holophyletic) domains.

Recognizing the archaeobacteria was a very important achievement that has stood the test of time; unfortunately, Woese has persistently misunderstood their evolutionary significance, the forces that generated them and their time of origin. It is quantum evolution during the relatively recent neomuran revolution and the immediately subsequent origins of archaeobacteria and eukaryotes (Cavalier-Smith, 1987a, b), not early divergence (Woese & Fox, 1977) and rampant lateral gene transfer (Woese, 1998, 2000), that is responsible for the sharpness of the boundaries between the three domains for many characters. The fact that many other features do not show such strongly marked differences is attributable to their relative stasis and the mosaic nature of evolution during the only partially revolutionary transitions between the three domains.

Rooting the tree of life and eubacterial megaevolution

It is more difficult to root trees than to work out affinities; it is easier to see that taxon A is more similar to taxon B than to C than to determine whether A and B are sisters, A is ancestral to B or B is ancestral to A, or whether instead A is actually cladistically closer to D than to any of A, B or C, but D's genealogical relationship to A is obscured through its greater divergence from the common ancestor.

Morphologically, the most fundamental dichotomy within bacteria is between bacteria bounded by one membrane [subkingdom Unibacteria (Cavalier-Smith, 1998) or subdomain Monodermata (Gupta, 1998b)] and those bounded by two concentric membranes [subkingdom Negibacteria (Cavalier-Smith, 1998) or subdomain Didermata (Gupta, 1998b)]. Table 1 summarized their classification. I have always considered that both groups are monophyletic (in the proper, classical non-Hennigian sense), but that Unibacteria are paraphyletic because eukaryotes also have only a single bounding membrane and almost certainly evolved from them. The Hsp60 tree shows a monophyletic Negibacteria and Posibacteria (Gupta, 1998a). Much more importantly, Gupta (1998a) has shown that several indels in proteins show that Unibacteria and Negibacteria are both monophyletic; the bacterial Hsp70 tree (Gupta *et al.*, 1999) also partitions between monophyletic Unibacteria and Negibacteria, but this fact would not be germane to the issue of the monophyly of Unibacteria if the archaeobacterial Hsp70 genes were derived secondarily from Posibacteria by lateral gene transfer, as I argued above is likely. Apart from misplacing *Thermotoga* slightly,

the concatenated RNA polymerase tree shows a monophyletic Unibacteria (Klenk *et al.*, 1999).

Given that Unibacteria are probably paraphyletic not holophyletic and the evidence that eubacterial radiation took the form of an almost irresolvable big bang, we should expect the demonstration of their monophyly (specifically paraphyly) to be relatively difficult. Were it not for the facts of quantum and mosaic evolution emphasized above, one would expect the relatively much more recent branching of the neomura within the actinobacteria to be resolved much more easily. Unfortunately, however, quantum evolution for all the characters listed in Table 2 is likely to be so extreme as to produce such excessively long branches that long-branch artefacts will cause them to branch near the base of the eubacteria rather than within Actinobacteria. The situation is similar to the gross problem in eukaryote rRNA trees (Fig. 6) that falsely put microsporidia near the base (Vossbrinck *et al.*, 1987) rather than in the correct, highly derived position within the fungi (Embley & Hirt, 1998; Keeling & McFadden, 1998; Roger, 1999; Hirt *et al.*, 1999; Keeling *et al.*, 2000; Van de Peer *et al.*, 2000; Cavalier-Smith, 2000c). The problem is much worse in bacteria, for two reasons. Firstly, it affects not just rRNA, but hundreds of proteins that almost certainly underwent quantum evolution almost simultaneously. Secondly, the metabolic proteins that did not undergo quantum evolution are the very proteins that on theoretical grounds (see discussion later) and according to empirical evidence (Rivera *et al.*, 1998; Jain *et al.*, 1999) are most prone to lateral gene transfer and also to multiple gene losses and are therefore likely to give thoroughly confusing trees. We are thus caught between the Scylla of quantum evolution and the Charybdis of lateral gene transfer. The fact that many genes also evolve too fast to be useful for deep phylogeny also greatly reduces the number of genes that might give useful phylogenies. A great deal of work will be necessary to see if we can sort out this confusion. Until this is done, we shall not know how much weight to give to the observation of Gupta (1998a) that 44 protein trees other than Hsp70 show the relationship between archaeobacteria and Posibacteria predicted by the neomuran theory, since, of course, some trees do not, as one also expects, given the demonstrated importance of quantum evolution and lateral gene transfer and differential losses of paralogues and the greater impact of tree reconstruction artefacts when taxa are sparsely sampled.

Despite these severe practical problems in testing the monophyly of Unibacteria other than by means of indels, which are very powerful, I consider that the distinction between Unibacteria and Negibacteria is as important as that between Eubacteria and Archaeobacteria for understanding cell evolution (Cavalier-Smith, 1987b, 1998). For different reasons, Blobel (1980) and Cavalier-Smith (1980) suggested that Negibacteria were ancestral and Posibacteria derived. Blobel (1980) postulated that Negibacteria were

Fig. 8. Hypothetical phylogeny of photosynthesis. The ancestral reaction centre was a homodimer with two bound quinones, each donating electrons to a primitive cytochrome *bc₁* complex (not shown). Gene duplication to form a heterodimer speeded transfer by passing electrons asymmetrically from M to L subunit quinone in green non-sulphur and purple bacteria. A common ancestor of cyanobacteria and heliobacteria formed two distinct homodimers from the L and M subunits, adding iron-sulphur clusters (F) to one; this was retained as a homodimer in heliobacteria but differentiated into the more complex, heterodimeric *psaA/B/C* photosystem I in cyanobacteria. The other homodimer was lost by heliobacteria but retained by cyanobacteria, where it associated with phycobilisomes and an oxygen evolution centre (Mn) to form photosystem II; asymmetric electron transfer between the two quinones was restored by gene duplication, yielding a D1/D2 heterodimer. Green sulphur bacteria also underwent homodimerization and Fe/S cluster addition, possibly independently.

formed by the 'gastrulation' of an 'inside-out-cell' and I postulated a mechanism for the loss of the outer membrane (murein hypertrophy to form the classical Gram-positives; see p. 906 of Cavalier-Smith, 1980). Because I favoured the view that the first cell was photosynthetic and thought that all bacterial photosynthesizers were Negibacteria, I developed Blobel's 'inside-out-cell' theory into a detailed explanation of the origin of the first cell, which I assumed to be a negibacterium, in particular, a photosynthetic green bacterium (Cavalier-Smith, 1985a, 1987a). The 'inside-out-cell' or obcell had the advantage over classical theories of circumventing the problem of the impermeability of simple lipid bilayers to nucleotides and amino acids and also of explaining the origin of the two negibacterial membranes.

The obcell theory has recently been greatly simplified and used to explain the early evolution of the genetic code (Cavalier-Smith, 2001). I have argued that a bioenergetic system using prebiotic high-energy inorganic oligophosphates and polyphosphates co-evolved with early genetic systems on the outer surface of an obcell to yield a genetic code for 10 prebiotic amino acids. Subsequent fusion of two cup-shaped obcells provides the first explicit gradual explanation of the origin of the first cell or protocell, which was bounded by an envelope of two membranes (Cavalier-Smith, 2001). The protocell is held to have successively evolved CO_2 fixation, photoreduction and soluble metabolism and expanded the genetic code to 22 amino acids. Thereafter, it increased its metabolic virtuosity and the complexity of its envelope, evolving peptidoglycan and lipoprotein to form the ancestral eubacterium (Cavalier-Smith, 2001).

The probable antiquity of green bacteria

Recent analyses of the evolution of photosynthesis give considerable support to the rooting of the tree among photosynthetic negibacteria. Gene-duplication trees of paralogous proteins involved in bacteriochlorophyll and chlorophyll synthesis suggest that the root of the photosynthetic eubacterial tree lies on one side or other of the green bacteria (Xiong *et al.*, 2000). All the different trees favoured one or other position but could not decide robustly between them. Xiong *et al.* (2000) suggested that the root was between the green bacteria and the purple bacteria (photosynthetic proteobacteria), as this was found more often than the alternative position between the green and purple bacteria on the one hand and the cyanobacteria and heliobacteria on the other. In my view, neither position is as likely as one within the green bacteria, precisely between the green sulphur bacteria and the green non-sulphur bacteria. Xiong *et al.* (2000) say that their trees disagree with the rRNA trees and attribute this to lateral transfer of chlorophyll biosynthesis genes, an interpretation echoed by Green (2001) and Blankenship (2001). I strongly doubt that lateral transfer is the explanation. I consider that the problem lies instead in misrooting of both the protein and the

rRNA trees caused by unequal evolutionary rates. If they were both rooted between the two groups of green bacteria, as I propose, they would actually be almost congruent. Fig. 8 summarizes a scheme for the evolution of the photosynthetic reaction centres that is simpler than other published schemes, yet the branching order between the phyla is the same as on the 16S rRNA tree (Hugenholtz *et al.*, 1998a, b).

I consider that the chlorosome must have been present in the common ancestor of Chlorobacteria and Sphingobacteria and lost by the members of both lineages that do not have it when they secondarily became non-photosynthetic. It is a unique structure, at least as complex as the phycobilisome of cyanobacteria. It has about 10 different proteins, at least two complex porphyrins (bacteriochlorophyll *c* and bacteriopheophytin) and carotenoids that have to co-operate in a complex. Its structural complexity and the physiological necessity for dependence on complex metabolism to make its porphyrin and carotenoid constituents means that it is no more likely to have been transferred laterally from green sulphur bacteria to green non-sulphur bacteria or the reverse than are ribosomes. The very robust clade comprising both green bacterial groups on gene trees for two sets of bacteriochlorophyll synthesis enzymes (Xiong *et al.*, 2000) is consistent with this, but does not distinguish between their holophyly or paraphyly. The facts that this grouping is so robust and the branches of each subgroup are so short mean only that the enzymes are very conservative and slowly evolving in the green bacteria; one does not need lateral gene transfer to 'explain' it. The greater distance from purple bacteria and the cyanobacterial/posibacterial clade implies that these genes underwent accelerated evolution in the ancestors of each of these two groups. This is hardly surprising, since they each evolved novel pigments: chlorophyll *a* in cyanobacteria and the related hydroxychlorophyll *a* and bacteriochlorophyll *g* in heliobacteria and bacteriochlorophyll *b* in purple bacteria. If my rooting of the tree is correct, they also independently lost chlorosomes and bacteriochlorophyll *c* and independently inserted their Mg-porphyrin antenna pigments into the cytoplasmic membrane instead.

This hypothesis (Fig. 8) is much simpler mechanistically than earlier ones that invoke fusion of organisms having different reaction-centre types (Blankenship, 1994) or lateral transfers of genes (Xiong *et al.*, 1998, 2000). It is also simpler than the idea of a complex ancestral type with two photosystems and differential loss in different lineages (Olson & Pierson, 1987). Compared with that complex scheme based on the earlier hypothesis of Granick (1965) that the cyanobacterial system is ancestral, it much better fits the sequence trees and the palaeontological evidence that early photosynthetic ecosystems were anaerobic plus its lack of evidence for cyanobacteria before 2.5 ± 0.3 Gy ago. This later origin of the oxygenic cyanobacteria and their posibacterial sisters

compared with the primary eubacterial radiation 3.5–3.7 Gy ago probably explains why most sequence trees group them consistently together but fail to reveal a robust branching order for the other three photosynthetic groups. Unless the deep divergence of the two green-bacterial classes on the rRNA tree and on trees for non-photosynthetic proteins is an artefact (conceivably caused by their thermophily), it is consistent with my thesis that the divergence of Eobacteria and Glycobacteria may have been the first bifurcation in the tree of life. Congruence between numerous protein trees and the rRNA tree is surely a useful criterion for correct rooting.

However, whichever position is correct for the root of the bacteriochlorophyll biosynthesis gene trees (Xiong *et al.*, 2000), the heliobacteria are in a derived position as sisters of cyanobacteria. This is also shown by the cytochrome *b* trees. In the best recent extensive eubacterial rRNA analysis (Hugenholtz *et al.*, 1998a), the only reasonably robust relationship between eubacterial phyla was that between cyanobacteria and Posibacteria, which include heliobacteria (Table 1). The recent discovery of heliobacteria with endospores (Ormerod *et al.*, 1996) supports the inclusion of heliobacteria within the Posibacteria, suggested both by the rRNA tree and by their single membrane, unlike all other photosynthetic eubacteria, which are negibacteria with two. A two-amino-acid insertion in pyruvate kinase implies that Endobacteria (endospore-forming low-G+C Gram-positives, within which *Heliobacterium* is nested unambiguously in Hsp70 trees; Gupta *et al.*, 1999) are holophyletic. These data collectively strongly support the view that Posibacteria were derived compared with negibacteria and that they therefore must have evolved by the loss of the outer membrane, as Blobel (1980) and I (Cavalier-Smith, 1980, 1987a, b) argued. This implies that the ancestral eubacterium was a negibacterium with an envelope of two membranes. The obcell theory simply explains how it evolved (Cavalier-Smith, 2001; Maynard Smith & Szathmáry, 1995).

Three other suggestions have been made as to how the outer membrane evolved, but none is very plausible. If posibacteria are in fact derived, as phylogeny indicates, none of these explanations, assuming the ancestral eubacterium to have been a posibacterium, is relevant. Dawes (1981) suggested that negibacteria evolved from an endobacterium and acquired their outer membrane by retaining the inner forespore membrane after spore germination. Chater (1992) suggested an alternative mode of origin of the double envelope by one hypha growing within another, as he has observed in *Streptomyces*; however, the signature sequences unique to Actinobacteria (high-G+C posibacteria; Table 1), notably a four amino acid insertion in DNA gyrase, imply that they are a uniquely derived eubacterial group (Gupta, 1998a), not ancestral to Negibacteria. To evolve an outer membrane required the insertion of porins and the development of focal adhesions (Bayer's patches) to allow lipids to move

from the cytoplasmic membrane during growth and the evolution of special periplasmic chaperone systems and secretion mechanisms. It could not have evolved in a sudden saltatory fashion, as these proposals assume. The discovery of four protein translocases in the outer membrane and three in the inner membrane (Stuart & Neupert, 2000) means that protein translocation is more complex in Negibacteria than in Posibacteria. Envelope evolution must have been gradual, over many generations; the obcell fusion theory (Cavalier-Smith, 2001) is the only one yet proposed that allows this via mechanistically plausible and arguably viable intermediates. The origins of these protein translocases must have been central to early negibacterial evolution. Rizzotti (2000) proposed a third, purely conjectural method of evolving a negibacterium from a posibacterium (type unspecified) involving protruding blebs, which seems even less plausible than the others.

I suggested earlier that the simplest teichoic acids found in most low-G + C posibacteria (glycerol phosphate co-polymers) might be exoskeletal remnants of a GNA world (preceding the RNA/protein world that, in turn, probably preceded the present DNA–RNA/protein world) where glycerol polynucleotides rather than RNA were the genetic material (Cavalier-Smith, 1987a). However, the evidence discussed here for a significantly later origin for Posibacteria than for Negibacteria rules this out. Teichoic acids (many more complex) may instead have been adaptations to help Gram-positive bacteria colonize terrestrial environments more readily by resisting drying in soils (Cavalier-Smith, 1980).

Compared with this strong evidence for the root among the Negibacteria, the common assumption that it is among the Unibacteria is almost devoid of support. Gupta (1998a) argues that Unibacteria are ancestral and Negibacteria are derived, because Negibacteria alone have an insertion in Hsp70 that is absent from a paralogue present in all organisms. However, the alignment of the paralogue in this area is very subjective and the argument not convincing.

Though I have severely criticized some widespread interpretations of certain features of rRNA trees, we must not throw the baby out with the bath water. rRNA trees do tell us something reliable! But we can only tell what that is by seeking features congruent with multiple protein trees and other evidence from cell biology and palaeontology. The sisterhood of archaeobacteria and eukaryotes and of cyanobacteria and posibacteria are two. So also is the sudden radiation of all the eubacterial phyla listed in Table 1. As argued above, this radiation – the eubacterial big bang – probably corresponds to the rapid radiation of bacterial photosynthesis in early Archaean microbial mats. Only two phyla (Spirochaetae and Planctobacteria) are entirely non-photosynthetic; if photosynthesis evolved in the protocell substantially prior to the cenancestor, as I have argued (Cavalier-Smith, 2001), their ancestors must have lost photosynthesis,

which has clearly happened several times within all other eubacterial phyla except Cyanobacteria. A deep divergence between the green sulphur and green non-sulphur bacteria is shown by both rRNA trees and several apparently reliable protein trees, such as Hsp70 and Hsp60. I take this as evidence that the green bacterial phenotype is very ancient and goes back to the time of the big bang itself. Protein trees also agree with rRNA trees in showing the monophyly of five eubacterial phyla, Posibacteria, Cyanobacteria, Spirochaetae, Proteobacteria and Sphingobacteria (green sulphurs and the Flavobacteria/*Cytophaga* lineages), so these taxa are well founded. But they only sometimes show the Eobacteria (green non-sulphurs, *Deinococcus* and *Thermus* lineages; Table 1) or Planctobacteria as poorly supported clades.

The simplicity of photosynthesis in heliobacteria, which have the simplest carotenoids (Takaichi *et al.*, 1997), is deceptive. Instead of being a precursor of the biochemically and ultrastructurally more complex systems in Negibacteria, it seems that it was simplified secondarily by loss of the chlorosomes.

The structure of the eubacterial tree

The sequence trees of the photosynthetic genes and of rRNA together provide a largely congruent and robust branching order for the five ancestrally photosynthetic eubacterial phyla. But where do the other two phyla, Planctobacteria and Spirochaetae, fit in? The 16S rRNA (Hugenholtz *et al.*, 1998a, b) and Hsp70 trees do not give robust positions for them, probably because they are part of the rapid early eubacterial radiation. In such cases, indels in proteins are sometimes very useful in grouping certain taxa (Gupta, 2000). Fig. 7 shows that several indels support the relationships argued above for the four photosynthetic phyla. Two single-amino-acid indels in very different, highly conserved proteins cleanly divide eubacteria into the same two groups: one in the division protein FtsZ and one in the chaperonin Hsp60 (Gupta *et al.*, 1999). Note that my interpretation of the Hsp60 indel differs from that of Gupta. Planctobacteria, Proteobacteria, Sphingobacteria and Spirochaetae all have a conserved asparagine at position 153; cyanobacteria all have a conserved glycine, whereas all the other groups have a ‘deletion’. He assumes that the asparagine and glycine are homologous and are ‘evidence’ that cyanobacteria are specifically related to the other four taxa. I disagree entirely. It is simpler to suppose that the ancestral eubacterium had no amino acid at that position (as in Eobacteria and Posibacteria) and that the glycine was inserted in the ancestral cyanobacterium and the asparagine in the common ancestor of the other four groups. The indel in FtsZ might also be an insertion in that same common ancestor or a deletion in the common ancestor of Cyanobacteria, Posibacteria and Eobacteria (if my rooting between Eobacteria and Glycobacteria is correct, it would be an insertion).

A third indel, of four amino acids in alanyl-tRNA synthetase, specifically groups Planctobacteria, Proteobacteria and Sphingobacteria to the exclusion of Spirochaetes. If the tree is rooted correctly then Planctobacteria, Proteobacteria and Sphingobacteria form a clade defined by this highly conserved four-amino-acid insertion. I suggest that Proteobacteria and Planctobacteria are sister phyla and group them as the superphylum Exoflagellata. Further data are needed to test this. As here constituted, Planctobacteria consist of the Planctomycetales and Chlamydiae with protein walls and the Verrucomicrobiae with either protein or peptidoglycan walls. Although they group together on some published trees, the evidence for monophyly of Planctobacteria is currently weak. I have grouped Chlamydiae and Planctomycetales together on the assumption that peptidoglycan was replaced by protein only once in their common ancestor (Cavalier-Smith, 1987a). However, although it would be reasonable to suggest that Chlamydiae evolved from endoparasitic, peptidoglycan-free Verrucomicrobiae, it is likely that Planctomycetales lost their murein independently from free-living ancestors. Although the alanyl-tRNA signature sequence (and their endocellular habit) make it almost certain that chlamydias are derived from ultimately free-living ancestors with peptidoglycan, the possibility that Planctomycetales might be primitively without murein cannot yet be ruled out. They deserve much more intensive molecular study to determine whether their unique features are derived or are the result of earlier divergence than I have assumed here. Unless the root of the tree really lies between Planctomycetales and all other bacteria, which is possible but unlikely, the ancestral eubacterium (the cenancestor of all life) would have had peptidoglycan. I have suggested that the origin of peptidoglycan significantly prior to the cenancestor should be taken as the boundary between protocells and stem eubacteria (Cavalier-Smith, 2001).

The above analysis indicates that, if we root the tree between Eobacteria and Sphingobacteria, we can construct a tree (Fig. 7) in which the branching order is congruent with the rRNA tree apart from its misplacement of the hyperthermophiles (Fig. 6), with the protein trees for Hsp60 and Hsp70, with the trees for photosynthesis-related proteins, with the indel data of Gupta *et al.* (1999) as here reinterpreted and with the evolution of ultrastructural features and chemical composition of the cell envelope and photosynthetic machinery of bacteria that I have particularly emphasized. The congruence of all these different lines of evidence suggests that Fig. 7 is an excellent working hypothesis for bacterial relationships. It gives a sensible picture of organismal phylogeny with no confusion at all from lateral gene transfer. The early bifurcation within Glycobacteria divides them into two major branches: Cyanobacteria/Posibacteria and Proteobacteria / Planctobacteria / Sphingobacteria / Spirochaetes, which I designate the CP and the PPSS

branches of the glycobacteria. Unlike the curious linear pattern of Gupta *et al.* (1999), reminiscent of the eighteenth-century ladder of life, this is a normal branched phylogeny, much easier to reconcile with conventional molecular trees and our general understanding of the divergent processes of evolution. Fig. 7 also differs profoundly from Gupta's scheme in accepting wholeheartedly the holophyly of archaeobacteria and in rooting the tree within the negibacteria, not the posibacteria. It is not set in stone and should be tested rigorously by other data.

Evolution of flagella, gliding motility and spirochaetes

A key question in bacterial evolution is when did flagella arise? Eobacteria, Sphingobacteria and Cyanobacteria lack flagella and often have gliding motility instead. All other bacterial phyla have flagella, though some subgroups within them have lost them secondarily. If cyanobacteria are sisters of Posibacteria, as the evidence discussed above indicates, the cyanobacterial cenancestor must have lost flagella. If the trees in Figs 1, 2 and 7 are rooted correctly, Cyanobacteria and Sphingobacteria lost flagella independently. If spirochaetes are placed correctly on Fig. 7, then the periplasmic location of their flagella is a secondary condition. If, contrary to my interpretation of the Hsp60 indel, the root of the tree should really be between spirochaetes and all other bacteria, which would be compatible with all other data, then their periplasmic flagella and normal external flagella might have diverged as alternatives at the very origin of flagella. I was once tempted by such a view (Cavalier-Smith, 1992b), as I then mistakenly thought that spirochaetes had no lipopolysaccharide and so might be primitive; though I now consider spirochaetes as derived, we need more substantial evidence to verify their position. Spirochaete flagellar shafts are markedly more complex than those in other bacteria, consisting of three different proteins surrounded by a fourth sheath protein (Li *et al.*, 2000). Although flagella have been lost repeatedly, they might have evolved after the cenancestor (Fig. 7; Cavalier-Smith, 1992b) rather than beforehand (Cavalier-Smith, 1987a).

The flagellar basal body and the need for its co-evolution with the cell wall are so complex that we can rule out lateral transfer of the whole apparatus, so the conclusion that the glycobacterial cenancestor had flagella is probably robust. To understand the evolution of gliding motility, we need to know whether it is homologous in cyanobacteria, Eobacteria and Sphingobacteria and whether its molecular basis is simple enough for it to be a possible candidate for lateral gene transfer. If it is homologous and was transmitted only vertically, it must either have been present in the cenancestor or, as is sometimes suggested, have actually have evolved from flagella

that lost their shafts; in that case, the driving mechanism could be homologous, but its use for gliding rather than swimming polyphyletic.

Eobacteria and the nature of the cenancestor

I postulated previously that the absence of lipopolysaccharide in Eobacteria is a primitive character (Cavalier-Smith, 1992b). It is a very complex molecule with highly elaborate biosynthesis and secretory requirements, so it may well have evolved after the cenancestor. As there seems no obvious reason why Eobacteria should have lost it if they ever had it (the highly reduced chlamydias have retained it), Fig. 7 assumes that they never had it. If eobacteria are indeed primitively without lipopolysaccharide, this makes them sisters to all other organisms. Thus, their lack of flagella may also be the primitive state. The exceptional radiation resistance of the eubacterium *Deinococcus* may be an ancestral character inherited from the Archaeon before the growth of the ozone layer. An RNA-binding protein that binds to several small RNAs seems to be involved in this resistance (Chen *et al.*, 2000). A third putatively primitive character is the absence of gas vesicles, which are found in all the other four bacterial phyla with photosynthetic members and in archaeobacteria. Although the clustered nature of gas vesicle genes might lend them to lateral gene transfer, there is no convincing evidence for this. Of the 14 halobacterial gas vesicle genes, homologues of all the eight essential ones are found in posibacteria (Offner *et al.*, 2000), their ancestral group, so inheritance could have been vertical.

Thus, I suggest that the cenancestor was an anaerobic, non-sulphur green bacterium with chlorosomes, bacteriochlorophyll *a* and *c*, carotenoids, peptidoglycan with ornithine but not diaminopimelic acid, RuBisCO and gliding motility but no flagella, lipopolysaccharide or gas vesicles. It is unclear whether Eobacteria are paraphyletic or holophyletic; on the rRNA trees of Barns *et al.* (1996) and Kyrpides & Olsen (1999) they are clearly holophyletic, whereas in Hugenholtz *et al.* (1998a) they are only barely together and, in some published trees, are not grouped at all.

The present re-rooting of the universal tree in conjunction with the generality of quantum evolution following gene duplication to form divergent paralogues makes it necessary to reevaluate conclusions about the nature of the cenancestor based on such trees. For example, it is unparsimonious to assume that the cenancestor had duplicates of both the ornithine carbamoyltransferases and the aspartate carbamoyltransferase genes and differential losses among bacteria (Labedan *et al.*, 1999). The paralogue tree is more simply explicable by a single cenancestral version of each, no differential loss, but an artefactual mid-point rooting of each paralogue by the other, plus a few lateral gene transfers.

RuBisCO: vertical and lateral evolution

As bacteria have several carbon-fixation enzymes, we do not know which came first. I once interpreted this diversity as evidence that autotrophy evolved polyphyletically after the cenancestor, suggesting that it was a heterotroph without RuBisCO (Cavalier-Smith, 1987a). If, as I argue, the root of the tree of life lies between the two groups of green bacteria, it is probable that the cenancestor had RuBisCO, as it has recently been found in a green non-sulphur bacterium (Ivanowsky *et al.*, 1999) and is found in purple bacteria that lie on one side and in cyanobacteria and posibacteria that lie on the other side of the major glycobacterial bifurcation shown in Figs 2 and 7. As there is evidence for a relatively recent lateral gene transfer of RuBisCO between these two clades (Paoli *et al.*, 1998; Horken & Tabita, 1999), RuBisCO phylogeny is less easy to treat cladistically than that of most of the other characters emphasized in this paper. If lateral transfer also occurred in its early evolution, the cladistic conclusion that it was in the ancestor need not be valid; however, we should bear in mind that lateral gene transfer by replacement of a functionally equivalent activity, as in this case, may be intrinsically easier than acquiring a new function. Therefore, the rampant lateral transfers of RuBisCO (Delwiche & Palmer, 1996) may simply be quasi-neutral substitutions of preexisting genes, not *de novo* acquisitions by lineages formerly lacking it; if so, cladistic reasoning about its origin would be valid despite them. Loss of RuBisCO has also probably occurred, as it is absent from the posibacterial heliobacteria, but was almost certainly present in the common ancestor of cyanobacteria and posibacteria. The facts that green sulphur bacteria fix CO₂ by a reductive TCA cycle and that most non-sulphur green bacteria use the hydroxypropionate cycle do not mean that their common ancestor could not fix CO₂. Since a potential for a reductive TCA cycle exists in all anaerobic photosynthesizers, it was also found in the cenancestor and became the carbon-fixation method of Chlorobea after their ancestors lost RuBisCO. The deep divergence between the proteobacterial and the cyanobacterial/posibacterial variants of type I RuBisCO is simplest to explain if the cenancestor already had RuBisCO, which the fossil record indicates is likely, and if it simply corresponds with the phyletic divergence of the CP and PPSS branches.

The relatively strong depletion of ¹³C compared with ¹²C back to about 3.5 Gy ago (a likely date for the cenancestor) is normally interpreted as evidence that RuBisCO has been the major carbon fixer ever since that period. The weaker depletion of ¹³C in organic carbon in the period 3.5–3.8 Gy ago is similar to that caused by the reductive TCA cycle of green sulphur bacteria or the propionate cycle of green non-sulphur bacteria. This weak depletion could therefore be caused by the relative importance of green non-sulphur eobacteria being greater than subsequently. However, it could also be caused by enrichment caused

by heating of these partially metamorphosed rocks (Strauss *et al.*, 1992; Schidlowski, 2001) altering the ratio produced by RuBisCO. Possibly, both RuBisCO and hydroxypropionate cycle enzymes were used by green non-sulphur bacteria during that period. Especially when metabolism was beginning and pathways inefficient, it could have been more advantageous to add a second carbon-fixing enzyme than to improve an existing one slightly. Just as early steam ships also used sails, so early photosynthetic protocells may have evolved multiple pathways of carbon fixation. The first RuBisCO was probably not the now widespread multi-subunit type I enzyme, but the simpler and smaller single polypeptide type II RuBisCO now found only in dinoflagellates and some proteobacteria. After type I RuBisCO evolved, both were retained by the cenancestor, but the primitive type II version was lost from the posibacterial/cyanobacterial lineage following the basic eubacterial bifurcation shown in Figs 2 and 7. Several differential losses of both types and some lateral transfers can together explain their present distribution.

Losses of glutaminyl- and asparaginyt-tRNAs

Glutamine and asparagine are unusual in being encoded in different ways in different organisms. In some cases, they have their own tRNAs like other amino acids, but in others, they do not and are made by respective enzymic modification (transamidation; Curnow *et al.*, 1997) of glutamic acid or aspartic acid already covalently attached to their cognate tRNAs. Most eubacteria and eukaryotes have a conventional asparaginyt-tRNA synthetase, whereas most archaeobacteria amidate aspartyl-tRNA instead. Clearly, given the eubacterial root to the tree, asparaginyt-tRNA synthetase, which had evolved prior to the cenancestor, has been frequently lost. It once seemed likely that the ancestral archaeobacterium lost this enzyme, but its discovery in *Pyrococcus* (euryarchaeote) and *Pyrobaculum* (crenarchaeote) and the grouping of their enzymes as sisters to the eukaryotic ones (Woese *et al.*, 2000) makes it likely that they have been lost more than once within both euryarchaeotes and crenarchaeotes. They have also been lost independently in *Aquifex*, *Thermotoga*, several proteobacteria, *Chlamydia* and the actinobacterium *Mycobacterium*. I suggest that the presence of the transamidation alternative for these two amino acids strongly predisposed bacteria to lose them quite rampantly.

Since glutamyl-tRNA synthetase is found in all organisms but glutaminyl-tRNA synthetase is found only in eukaryotes and a few eubacteria (Proteobacteria, *Deinococcus* and *Porphyromonas*), it has been suggested that glutaminyl-tRNA synthetase evolved only in eukaryotes and was transferred laterally several times to eubacteria (Lamour *et al.*, 1994). However, recent phylogenetic analysis does not support this; the eubacterial sequences are not nested within the

eukaryotic ones, but are their sisters (Brown & Doolittle, 1999; Handy & Doolittle, 1999). The fact that both molecules are found widely in the Chromatibacteria (sulphur purple bacteria and their colourless descendants or β - and γ -proteobacteria, a clearly holophyletic group) makes it highly probable that both were present in their common ancestor. Inspection of numerous molecular trees suggests that Chromatibacteria are comparable in age to the α -proteobacteria, the ancestors of mitochondria, and about two-thirds the age of eubacteria as a whole. If eubacteria are 3.7 Gy old, Chromatibacteria would be about 2.5 Gy old, about three times the age of eukaryotes (0.85 Gy; Cavalier-Smith, 2002). Therefore, the glutaminyl-tRNA synthetase gene cannot have been transferred from eukaryotes to Chromatibacteria; successive transfers via *Deinococcus* proposed by Handy & Doolittle (1999) are even more improbable.

Multiple losses of enzymes are probably easier and more frequent than lateral gene transfers, contrary to recently fashionable assumptions. Since the *Porphyromonas* gene is as divergent as the chromatibacterial genes and one of the two *Deinococcus* genes very much more so, it is most likely that the ancestral eubacterium had both genes and that the glutaminyl-tRNA synthetase gene has been lost by eubacteria that do not have it. To explain why the eukaryote glutamyl-tRNA synthetases are so much more similar to eubacterial glutaminyl-tRNA synthetases than to their glutamyl-tRNA synthetases, we must suppose that the glutaminyl-tRNA synthetase gene underwent gene duplication in an ancestor of eukaryotes and one copy took over the glutamyl charging function, allowing the original glutamyl-tRNA synthetase gene to be lost. It is less easy to determine the ancestry of the archaeobacterial glutamyl-tRNA synthetases, since they do not branch within either eubacterial clade, but are somewhat more similar to the glutaminyl-tRNA synthetase genes. I favour the view that the postulated glutaminyl-tRNA synthetase gene duplication, reassignment of amino acid and loss of the eubacterial glutamyl-tRNA synthetases took place not in the ancestral eukaryote but in the neomuran common ancestor. Archaeobacteria then lost the glutaminyl-tRNA synthetase enzyme and the remaining enzyme diverged rapidly from its eukaryotic sister prior to the primary radiation of archaeobacteria and then evolved more slowly; such rapid early divergence could account for its not branching with its putative eukaryotic sisters. Thus, one gene duplication, one functional reassignment and two gene losses can account for the puzzling phylogeny and distribution of these enzymes. If my arguments are correct, lateral transfer cannot. This interpretation is consistent with the fact that glutamine synthetase is found throughout eubacteria. Therefore one cannot rationalize the frequent absence of the glutaminyl-tRNA synthetase by saying that they had not yet evolved glutamine synthetase and so did not need it. Type I glutamine synthetase is found in all

bacteria and probably evolved in the protocell. Type II glutamine synthetase probably evolved in actinobacteria, so both were present in the neomuran ancestor; differential loss of type I in the ancestral eukaryote and type II in the ancestral archaeobacterium explains their present distribution.

The fact that Actinobacteria share a type I- β glutamine synthetase with a 25-amino-acid insertion and regulation by reversible adenylation uniquely with Negibacteria suggests that this was the ancestral state; glutamine synthetase I- α is found instead in Endobacteria (including *Thermotoga*, further support for its placement therein) and archaeobacteria (Brown & Doolittle, 1997); this suggests either that glutamine synthetase I- α arose in the ancestral endobacterium and was transferred laterally to archaeobacteria or that it evolved in the ancestral posibacterium, co-existed for a period with type I- β in actinobacteria and was then lost by them but persisted in the lineage that gave rise to archaeobacteria. It is much more difficult than is often thought to distinguish between lateral transfer and multiple losses.

Lateral gene transfer and hyperthermophily

A quarter of a century ago, I thought that lateral gene transfer might be commoner than was then assumed (Cavalier-Smith, 1977). Now I think its frequency is often exaggerated. If we not only ignore quantum evolution by assuming that all genes are clock-like but also root the universal tree in the wrong place, we shall often be driven to invoke immensely more lateral gene transfer than really occurred. When these factors are taken properly into account, lateral transfer will be found to be much less frequent than many recent papers assert. Mosaic evolution with extreme *trans*-lineage rate variation is really the norm in large-scale protein evolution; the mythical molecular clock has never been demonstrated objectively to apply universally to any molecule, yet belief in its validity and overconfidence in and repeated dogmas about tree-rooting lie behind many recent assertions of rampant gene transfer based simply on statistical treatments of overall similarity. The idea that a cell or an organism is a mosaic of genes evolving in radically different temporal patterns can, when coupled with a judicious rooting of the tree of life, account for much of the superficially confusing pattern of gene distribution among bacteria without invoking the massive amounts of lateral gene transfer favoured by many recent authors.

However, there is enough good evidence for lateral gene transfer to indicate that it is a pervasive influence on bacterial evolution, and I do not deny its importance (Doolittle, 1999a, b, 2000) or ignore the difficulties it poses for reconstructing bacterial evolution (Doolittle, 1999a, b, 2000). However, Kyrpides & Olsen (1999) and Logsdon & Faguy (1999) point out correctly that vertical inheritance, plus differential losses and differential rates of change, are often

overlooked by suggestions like those of Nelson *et al.* (1999) and Aravind *et al.* (1998) of massive lateral gene transfer based simply on overall similarity. Proper phylogenetic analysis is needed to demonstrate lateral transfer, including correctly rooted trees with the right topology. Unfortunately, the rRNA tree of Kyrpides & Olsen (1999) does not meet that requirement. Much of the branching order of the eukaryote part of that rRNA tree is certainly wrong (see Fig. 6 and Cavalier-Smith, 2002), the eubacterial part is incorrect in at least three respects, notably the positions of *Thermotoga* and *Aquifex* (Fig. 6), and the position of the root is wrong. This makes it likely that Aravind *et al.* (1999) are partially correct and that some genes probably were transferred laterally between archaeobacteria and *Aquifex* and *Thermotoga*. Acquisition of genes specifically concerned in hyperthermophily by *Thermotoga* and *Aquifex* could have had great selective advantage by allowing them to colonize superhot environments. Thus, such transfer is evolutionarily plausible.

However, plausibility does not mean that it actually took place. We need proper phylogenetic analysis of each gene to assess whether or not it was transferred, like that of Nesbø *et al.* (2001), who clearly demonstrate transfers of two metabolic genes from archaeobacteria to *Thermotoga*. Four independent transfers of glutamate synthase from euryarchaeotes to *Thermotoga*, the halorespiring chlorobacterium *Dehalococcoides*, the posibacterium *Clostridium difficile* and the proteobacterium *Sinorhizobium* are all convincing. Three other transfers within the eubacteria involving proteobacteria are possible, but need many more data for other taxa to become convincing, since there is clear evidence also for paralogy and differential gene loss and also for poor resolution of the trees within eubacteria that might account for some of them. For the transfers from Archaeobacteria, the conclusions from the tree are firmly supported by the specifically archaeobacterial splitting of the protein into three separate genes (Nesbø *et al.*, 2001). These glutamate synthase transfers are simple gene replacements and might be neutral changes of no functional significance. By contrast, the acquisition by *Thermotoga* of the archaeobacterial myoinositol IP synthase gene (*ino1*) probably helped adapt it to high temperature and high salt by enabling it to produce the osmolyte di-myoinositol 1,1'-phosphate (DIP). The archaeobacterial origin of the *Thermotoga ino1* gene is supported strongly by the tree and also by the presence of flanking archaeobacteria-like genes (Nesbø *et al.*, 2001). However, I disagree with their suggestion of four other lateral transfers for this gene, which I think was led astray by their acceptance of the Iwabe *et al.* (1989) misrooting of the tree. All except two of the eubacterial genes in their prokaryote groups 2 and 3 are actinobacterial; I think that actinobacterial genes were probably vertically ancestral to the archaeobacterial and eukaryotic genes. I suggest that the *ino1* gene and the osmolyte originated in thermophilic actinobacteria

as an adaptation to thermophily. The existence of four major clusters separated by immensely long stems shows that quantum evolution has repeatedly distorted the *inol* tree, making its rooting problematic. The fact that *Streptomyces coelicolor* has three genes in two different major clusters is indicative of gene duplication and deep paralogy within actinobacteria. If the neomuran ancestor had several deeply paralogous genes, their differential survival in different lineages, rather than lateral gene transfer from archaeobacteria to actinobacteria, could account for the complex neomuran tree. However, I agree that lateral gene transfers to *Aquifex* and *Dehalococcoides* are needed to explain why they are the only negibacteria to have the gene, but I suggest that the donors were actinobacteria, not archaeobacteria. This interpretation better explains the distorted dimensions of the tree and involves two fewer transfers than that of Nesbø *et al.* (2001).

Aravind *et al.* (1998, 1999) assume that interdomain transfers involving the hyperthermophilic eubacteria were all from archaeobacteria. This assumption rests primarily on the false belief that archaeobacteria are ancient. If archaeobacteria are actually four times younger than eubacteria, then hyperthermophily might have first evolved in eubacteria, either in Thermotogales or Aquificales. Judging from the depth of their internal branches on rRNA trees, admittedly hazardous, both groups might be over half as old as the eubacterial ancestor and therefore twice as old as archaeobacteria. Possibly, some genes assumed to have moved from archaeobacteria to them might actually have evolved in a eubacterial hyperthermophile and moved later into the common ancestor of archaeobacteria. It is also possible that Thermotogales and Aquificales donated hyperthermophilic genes to archaeobacteria and to each other. The argument that reverse gyrase in both eubacteria is closely linked to 'archaeobacteria-like genes' (Forterre *et al.*, 2000) is unfortunately somewhat ambiguous if those adjacent genes also entered the archaeobacterial ancestor from a eubacterial hyperthermophile. The placement of the *Thermotoga* sequence among crenarchaeotes and the *Aquifex* one among euryarchaeotes (Forterre *et al.*, 2000) fits two independent lateral transfers from archaeobacteria; if the direction was the reverse, the eubacterial genes ought instead to lie between crenarchaeotes and euryarchaeotes if the conventional rooting of the archaeobacterial tree is correct. However, though suggestive, this evidence for lateral transfer from archaeobacteria is not yet compelling, because of low bootstrap support and limited taxon sampling. It will be important to repeat it with many more sequences.

Lateral transfers from eubacterial thermophiles might have played a part in the origin of hyperthermophily in the ancestral archaeobacterium, e.g. in the acquisition of sulphur-reducing enzymes as discussed above. Careful phylogenetic analyses are needed to see how many such transfers are likely and if their direction can be established. Some aminoacyl-tRNA synthetases

group *Aquifex* and *Thermotoga* together (but not with neomura), whereas others show them in their probably correct positions in Proteobacteria and Posibacteria, respectively (Woese *et al.*, 2000); the former were possibly transferred from one to the other. My hunch is that the depth of both Thermotogales and Aquificales on rRNA trees is greatly exaggerated and that hyperthermophily probably first evolved about 850 My ago in the ancestral archaeobacterium, so before then there were plenty of eubacterial thermophiles but no hyperthermophiles.

Recent worries that lateral transfer is so rampant (Doolittle, 1999a, b) that we may never reconstruct organismal trees are certainly false for eukaryotes and probably incorrect for bacteria. I agree with Doolittle (2000) that the widely accepted tree needs uprooting, not because of lateral transfer, which is not seriously confusing with respect to the root, but because quantum evolution caused misrooting of the paralogue tree. We may safely replant it as shown in Figs 1, 2 and 7. The idea that genome composition in the ancestor was so fluid (Woese, 1998, 2000) that we cannot use cladistic arguments to reconstruct it is more profoundly mistaken, being based on the basic misinterpretations of the universal tree and the evolutionary significance and timing of the differences between eubacteria and neomura explained above. It has long been clear (Cavalier-Smith, 1981, 1987a, b, 2001) that the ancestor was a normal eubacterium, not a progenote. As Woese (1998, 2000) and Doolittle recognize, the most readily transferred genes are not a random sample of the whole, but obey certain rules (Rivera *et al.*, 1998; Jain *et al.*, 1999; Martin, 1999). Two categories of protein must be most prone to lateral transfer. (i) Firstly, those that interact little if at all with other proteins and function on their own or by interactions with widespread small molecules in generalized ways that are not taxon-specific (for example many enzymes of glycolysis, aminoacyl-tRNA synthetases, glutamate synthase); from a functional viewpoint, endogenous proteins may be substituted by foreign ones with no great impact on fitness. Because of the neutrality of such substitutions, they may occur at rates dependent on the frequency and efficiency of uptake of foreign DNA and genetic drift. (ii) Secondly, those where acquisition of a single gene or operon may drastically improve fitness in a particular environment (e.g. the ability to degrade a natural biocide like penicillin or make an osmolyte like DIP, the ability to digest a novel food, e.g. cellulase, or the ability to bind to host cells or to fool host defences). At the opposite extreme, strongly interactive macromolecules that interact by binding to numerous other disparate cellular macromolecules will scarcely ever be subject to lateral gene transfer, though they have been transferred as part of a functioning macromolecular complex during the seven known examples of cellular symbiogenesis.

One class of genes particularly prone to lateral gene transfers is the aminoacyl-tRNA synthetases

(Doolittle & Handy, 1998). Yet, even here, as Woese *et al.* (2000) stress, transfers are relatively few and can themselves sometimes be used as important cladistic (shared transferred) characters that can be used to cement one part of the organismal phylogeny more firmly and (as I showed above) to reveal the relative timing of the origin of groups without fossils. Even in bacteria, where lateral transfer is undoubtedly commoner than in eukaryotes, it would be unwise to make lateral transfer the null hypothesis, as W. F. Doolittle at times almost seems to advocate. Technical artefacts are known to be rampant in tree construction and are inherently more likely than lateral gene transfer to be the cause of discordant trees. I am unconvinced, for example, by the recent claim for horizontal transfer of catalase peroxidase genes (Faguy & Doolittle, 2000), which could easily be such an artefact. Lateral transfer has been important in intron evolution; coupled with our ignorance of many important eubacterial lineages, this makes it difficult to determine when group I and group II introns originated. Some group I introns of purple bacteria appear to have originated from those of cyanobacteria by lateral transfer (Paquin *et al.*, 1999).

The ability to take up foreign DNA was probably present in the cenancestor, as the basic DNA-uptake machinery used for genetic transformation is found in all eubacterial phyla (Dubnau, 1999) and homologues of ComEC, the putative channel protein, occur in all eubacterial phyla including Eobacteria. The original function of this DNA-uptake machinery was probably trophic (Redfield, 1993) not genetic. Using foreign genes as food must date back to pre-cellular evolution (Cavalier-Smith, 2001).

Importance of gene losses in evolution

For over a century, multiple character losses have been a major problem for phylogenetic reconstruction that often leads a simple cladistic approach astray. With respect to losses, evolution is most certainly not parsimonious. One reason why unweighted parsimony is a philosophically and empirically unsound approach to phylogenetic construction is that losses and gains do not have equal weight. For vertical inheritance, the origin of complex characters (e.g. eyes, legs, wings, tails, cilia, bacterial flagella, photosynthesis) is orders of magnitude more difficult than their loss. When lateral transfer by cellular symbiosis became the popular explanation for the origin of mitochondria and chloroplasts, some enthusiasts for such lateral organelle transfers blithely postulated dozens of such origins. I have suffered decades of often dogmatic opposition to my arguments that such transfers are actually evolutionarily very difficult and that loss is far easier, but we now know that losses of mitochondria and chloroplasts are an order of magnitude more common than their gains (Roger, 1999; Cavalier-Smith, 2000a, b). Enthusiasts for lateral gene transfer are now making an analogous mistake, by overesti-

imating its frequency and underestimating the much higher frequency of gene losses.

Aravind *et al.* (2000) have shown that the lineage represented by the yeast *Saccharomyces cerevisiae* has lost 300 genes compared with *Schizosaccharomyces pombe* and other eukaryote outgroups, yet lateral gene transfers into or out of this lineage are unknown. Lwoff (1944) long ago stressed the importance of losses in biochemical evolution and decades ago it was the standard explanation for the highly variable enzymic capabilities of many bacteria. Gene duplication and differential loss of paralogues is very common in vertebrates (Page, 2000). It is probably also common in bacteria, as Martin (e.g. Nowitski *et al.*, 1998) has argued repeatedly. Doolittle (1999b) says that invoking paralogy and multiple losses 'can seriously violate the rules of parsimony'. But the 'rules' of unweighted parsimony ought to be violated, as they are philosophically and empirically wrong when we are comparing gains and losses. Weighted parsimony is sensible. Unweighted parsimony, merely comparing the numbers of losses and gains, is stupid. I predict that, when careful studies are done, gene losses will be found to be at least two orders of magnitude more common than lateral gene transfer in eukaryotes (excluding the very rare special case of cellular symbiogenesis that can simultaneously implant thousands of genes) and probably an order of magnitude more common in bacteria. If lateral transfer took hold as a null hypothesis or dogma, rather than as one of several possible explanations for conflicting trees, studies to test this would be impeded.

Consider the case of the evolution of isoprenoid biosynthesis in bacteria, recently well reviewed from a lateral gene transfer perspective (Boucher & Doolittle, 2000). Two different non-homologous multienzyme pathways are very widespread in bacteria. Both the mevalonate and the deoxyxylulose phosphate (DOXP) pathways are found in Eobacteria, Proteobacteria, Spirochaetes, Sphingobacteria and Posibacteria, but only the DOXP pathway is known so far from Planctobacteria and Cyanobacteria. Archaeobacteria and eukaryotes use the mevalonate pathway only, except that the DOXP pathway is also present in chloroplasts (encoded by nuclear genes). Given the trees of Figs 2 and 7, this distribution has a very simple explanation. Both pathways were present in the cenancestor; the DOXP pathway was lost in the neomuran cenancestor, but reacquired by plants through the symbiogenetic origin of chloroplasts, whereas the mevalonate pathway was lost instead by Cyanobacteria and Planctobacteria. Since the two pathways are distributed patchily within the five phyla that have both, there must also have been additional complete or partial differential losses of these enzymes within each phylum. There is no reason to think that either pathway was transferred laterally as a whole at any time, except by the symbiotic origin of plastids. In particular, there is no evidence whatever that the two pathways evolved in different groups.

Boucher & Doolittle (2000) do, however, present evidence for homologous gene replacement within the mevalonate pathway itself for HMG-CoA reductase, which makes the mevalonate. All HMG-CoA reductases are related, but form two sharply distinct clusters on trees. All neomuran enzymes are class 1, except those of the eukaryote *Giardia* and the archaeobacterium *Archaeoglobus*, which are class 2 enzymes. In the present state of knowledge, the suggestion that these two sequences were acquired by lateral gene transfer from eubacteria (Boucher & Doolittle, 2000) is rather convincing. But I do not agree with their conclusion that lateral gene transfer also occurred between eubacteria. They suggest this simply because some eubacterial enzymes are class 1 and some are class 2 enzymes. They assume that eubacteria originally had only class 1 enzymes and that the class 1 enzymes of *Streptomyces* and *Vibrio* were acquired from archaeobacteria by lateral gene transfer. The weakness of this interpretation is that they make no explicit suggestion about how and in what organism the difference between the class 1 and class 2 enzymes came about in the first place. Since Proteobacteria and Posibacteria can both have either class 1 or class 2 enzymes, they could have arisen by gene duplication in the ancestor, undergone early quantum divergence within the ancestral lineage and co-existed for substantial periods as eubacteria diversified, but were eventually lost differentially from many lineages. Contrary to what is asserted, the presence of a shared four-amino-acid insertion does not 'unequivocally support an archaeal origin'; that insertion could have been the ancestral state for the class 1 enzymes and have been lost in the ancestral eukaryote. Far from being unambiguous examples of lateral gene transfer, as claimed, the *Vibrio* and *Streptomyces* genes can easily be interpreted as vertically inherited with quantum divergence of paralogues and multiple losses. With the limited data, we cannot say which interpretation is correct.

Whether there were four or, as I suspect, only two cases of gene replacement of HMG-CoA reductase genes, replacing such a gene – or an aminoacyl-tRNA synthetase – by a functionally equivalent one would seldom have a big impact on organismal evolution. It is a bit like replacing a decayed timber in a historic building; even if the new beam comes from a different source, it does not alter the function or architecture of the building. Even if, over the centuries, every beam and brick were to be replaced, it would still be architecturally the same building and clearly distinct from others of different design that might similarly have been repaired with modern bricks. The form of the buildings does not depend on the 'phylogenetic source' of their building blocks. For understanding the evolution of metabolism, the pathway as a whole matters much more than the phylogenetic source of an individual enzyme. One could conserve the pathway without conserving any genes. It is a fallacy to suppose that reconstructing organismal phylogeny will only

be possible if there is a core of genes that are never transferred (Doolittle, 1999a, b). Even if all were transferred now and again, we could still construct good phylogenies of higher-level biological organization, since the same gene would not be being transferred in every lineage at the same time. There is no simple mapping between genes and organisms. For organismal evolution, what matters especially is the form of the organism, which is maintained by only a subset of genes. None of these has yet been shown to undergo lateral gene transfer.

Gene losses seem to have been frequent during the origin of archaeobacteria; apart from the Hsp90 shown in Table 2, Gupta (1998a) lists several dozen others. As mentioned above, hundreds more are likely. Crenarchaeotes probably lost even more genes than euryarchaeotes; in addition to losing histones, they lost the FtsZ and MinD division proteins. It will be interesting to compare their division mechanisms with those of the posibacterium *Ureaplasma* and chlamydias, which lost them independently (Bernander, 2000), presumably after their ancestors also lost peptidoglycan. It seems that the loss of murein predisposes bacteria to lose FtsZ, but not inevitably. One wonders whether some of these genes might have been lost because they could not adapt satisfactorily to the even higher temperatures favoured by crenarchaeotes compared with euryarchaeotes.

Putting the organism back into the picture

A moderate degree of lateral transfer does not threaten the enterprise of phylogenetic reconstruction of organismal evolution fundamentally, as Woese (2000) rightly argues; it merely adds an additional objection to the naïve view that we can rely on single-gene trees to do this and to the false view of an organism as merely the sum of its genes. To be dismissive of morphology and phylogenetic evidence from sources other than sequences (Woese, 1994, 1998) is a mistake. Even to understand the evolution of rRNA, for example, we need to understand how it has been influenced by the rest of the cell, e.g. the changes in the SRP in the stem neomuran, the need to transport ribosomal subunits across the nuclear envelope or the co-evolutionary impact of mitochondrial ribosomal proteins in the same compartments as cytosolic ribosomes. Although it is welcome that Woese (2000) now urges the importance of understanding 'cellular design' and the 'cellular fabric', his writings have persistently overlooked the most important features of 'cellular design', the key integrative importance of membranes in cell and organismal biology (for a simple summary of this, see Cavalier-Smith, 2001) and the role of cell skeletons. They therefore never came to grips with the key differences in cellular fabric and 'design' between negibacteria and unibacteria or with those between the ribosome-related secretory mechanisms of archaeobacteria and their posibacterial ancestors. There is much more to cells than translation,

transcription and replication, about the only features considered in the vague discussions of the progenote; we can reconstruct the ancestral cell much more concretely than that by using knowledge of cell biology (Cavalier-Smith, 2001).

Molecular biologists too often neglect the part that cell biology can play in understanding bacterial as well as eukaryotic evolution. Cells and organisms are composed of interdependent parts that never evolve entirely independently. Most of those that have been helpful in reconstructing organismal evolution (e.g. rRNA, tubulins, actin, RNA polymerase, protein synthesis elongation factors, cytochrome oxidase and cytochrome *b*) are strongly interactive and are thus strongly influenced by co-evolutionary forces. To make more sense of molecular evolution, we need much more emphasis on cell evolution. Deeper understanding of bacterial cell biology, still poorly known despite the sequencing of several genomes, will allow us to construct organismal phylogeny more satisfactorily than can statistical treatments of randomly chosen genes, which is even more likely to be frustrating in prokaryotes, where a higher proportion of genes encode general metabolic enzymes subject to lateral transfer, than in the structurally more complex eukaryotes. Genes that help define core bacterial morphology, like the seven protein translocases of the negibacterial double envelope, the determinants of the periplasmic location of spirochaete flagella and the discreteness of cyanobacterial thylakoids and the attachment of DNA and ribosomes to membranes, will enable us to reconstruct the organismal evolution of bacteria even though, within this solid framework, many soluble metabolic enzymes may ebb and flow from foreign sources.

This paper has attempted to show that one can reconstruct an organismal phylogeny for bacteria by integrating important morphological characters, indels and sequence trees. Certainly, the frequency of lateral transfers means that we cannot confidently construct organismal trees from single-gene trees. But no sensible person ever thought we could. There are several other reasons why reliance on single-gene trees was always naïve, e.g. differential and constantly shifting rates of change (between taxa and at different positions on a molecule), gene composition biases and shifts in covariations. It also means, as Doolittle (1999a, b) rightly points out, that we cannot simply sum up the changes in all genes and assume that this would give us the correct phylogeny. But, again, no sensible phylogeneticist ever thought it would. Phylogeneticists have always known that different rates and degrees of change, multiple losses and multiple convergences and parallelisms are phylogenetically confusing and often lead to error. But lateral gene transfer is only a special kind of convergence that can, in principle, be handled in the same way as was always done long before sequences entered the phylogenetic scene. The key principles are to look at all the evidence from all sources and to weigh it differentially according

to our knowledge of the organisms in question and general understanding of biology and evolutionary processes. There is no simple recipe for this that you can feed into a computer. You have to think and build on experience by trial and error. From centuries of experience, systematists know that reliance on one line of evidence is dangerous and that one cannot make simple a-priori rules that will always give a sound conclusion (Mayr & Ashlock, 1969). If these principles apply to phylogeny, they do so even more to taxonomy, where we are concerned not with deducing accurate trees but with using them, together with all other available knowledge, to place organisms into evolutionarily reasonable and useful groups. Classification is concerned with three things: grouping, naming and ranking. For brief statements of some philosophical principles underlying these, see Cavalier-Smith (1998) and Mayr (1998).

Bacterial megaclassification

In recent years, the higher-level classification of bacteria has become confused and unnecessarily complex through lack of attention to the principles of ranking and overemphasis on rRNA similarity as a single arbitrary criterion of relatedness. Ranking has become very unbalanced, with frequent mention of numerous bacterial kingdoms but almost no attempt to define bacterial classes comprehensively, apart from Cavalier-Smith (1992b). To impose more order on the vastly increased numbers of bacterial taxa up to the rank of order, we urgently need to group them into a reasonably limited number of classes that are organismally relatively homogeneous yet phylogenetically sound. I hope that most of the 29 classes recognized in the present system will be found useful and that they will exemplify a degree of organismal similarity appropriate for a bacterial class. By using classes in a more balanced way, as is customary for eukaryotes, we can reverse the recent unhelpful hyperinflation in the number of bacterial divisions and 'kingdoms'. The present bacterial classification is revised from earlier attempts (Cavalier-Smith, 1987a, b, 1991a, b, 1992b, 1998); I refer the reader to Cavalier-Smith (1998) for a general discussion of some of the principles of higher-level classification as applied to bacteria. Most groups labelled 'new' in Table 1 were actually proposed with the same name (and similar or identical circumscription) in one of these earlier publications, but, as they were in the 'wrong' journal, I now validate them by designating them as new and providing fresh diagnoses for this 'official' journal; for historical continuity, Table 1 cites my publications where the names were first published. In some cases, I have changed somewhat the rank of taxa suggested by myself or others, either to allow probably related taxa to be grouped together more easily or to simplify the classification. I am confident in the monophyly of six of the eight phyla; but that of Planctobacteria and my inclusion of Ferrobacteria within the Proteobacteria need to be tested rigorously

by numerous good protein trees and indel data. Planctobacteria form a clade on some rRNA trees (e.g. Dojka *et al.*, 2000) but not on others (Hugenholtz *et al.*, 1998a, b; Ward *et al.*, 2000), but it seems premature to conclude that they are not monophyletic. Posibacteria are undoubtedly paraphyletic, because of their neomuran descendants, and Cyanobacteria and Proteobacteria are technically so because of their chloroplast and mitochondrial descendants. Eobacteria might be holophyletic or paraphyletic.

The present system recognizes only eight bacterial phyla, not 10 as described previously (Cavalier-Smith, 1998). Placement of Heliobacteria within Posibacteria, as an order rather than class (Cavalier-Smith, 1991a) or a separate phylum (Cavalier-Smith, 1998), is now supported by signature sequences and the Hsp70 tree (Gupta *et al.*, 1998a), as it always was by rRNA (Woese, 1987); this reduces the number of eubacterial phyla to seven, if we also set aside Eurybacteria (Cavalier-Smith, 1998) as probably polyphyletic.

We must remember that our goal is to classify organisms in a way that is consistent with phylogeny but places boundaries between groups at points of maximal phenotypic discontinuity. Though I earlier ranked archaeobacteria as a subkingdom (Cavalier-Smith, 1983b) or infrakingdom (Cavalier-Smith, 1998), this is no longer necessary. Archaeobacteria are just a fascinating unbacterial phylum specialized for hyperthermophily, which share numerous distinctive features with eukaryotes, but are insufficiently diverse to be subdivided into phyla. It is a pity that the name Metabacteria (Hori *et al.*, 1982) did not catch on for archaeobacteria, since they are undoubtedly the most derived and recent of all bacterial phyla, as several scientists have long argued (Van Valen & Maiorana, 1980; Hori *et al.*, 1982; Cavalier-Smith, 1987c; Forterre, 1996; Gupta, 1998b) and are certainly not a 'primary line of descent' (Pace *et al.*, 1986). Subdivision rank (the same as for vertebrates or seed plants) for Euryarchaeota and Crenarchaeota (Cavalier-Smith, 1998) is amply sufficient; this possibly separates them at even too high a rank. Eurythermea and crenarchaeotes are basically similar in cell structure and physiology, differing mainly in losses by the crenarchaeotes (e.g. in FtsZ and histones) rather than in major innovations by either. Phenotypically, there is much to be said for my earlier inclusion of both in a single group, Sulfobacteria (Cavalier-Smith, 1986). But for the fact that one methanogen has retained sulphur reduction, I would be tempted to retain the taxon Sulfobacteria for Crenarchaeota plus Eurythermea. Instead, I suggest retaining the term sulfobacteria (lower case) as a useful physiological and ecological descriptor for all sulphur-reducing archaeobacteria – the ancestral archaeobacterial organizational grade.

I anticipate that the recently cultivated, hyperthermophilic Korarchaeota (Barns *et al.*, 1996) will probably also turn out to be sulphur-dependent. I place them as an unranked group within archaeobacteria; from rRNA

trees, it is unclear whether they branch within Crenarchaeota (likely) or are an outgroup to all other archaeobacteria (Barns *et al.*, 1996); they should be examined for histone and FtsZ genes to help establish their position. Placing all crenarchaeotes in one class is consistent with their basic organismal similarity as currently understood; when the phenotypes of the psychrophilic Cenarchaeales are better known, their inclusion in the same class might need revision, but the fact that they have tetraether lipids like other crenarchaeotes (DeLong *et al.*, 1998) suggests that they may not be radically different, apart from not being hyperthermophiles. Euryarchaeotes are organismally much more diverse and merit five classes; I group the three derived, typically non-sulfobacterial ones as a new superclass Neobacteria; the name indicates that they were probably the latest of all bacterial supra-class-level taxa to evolve. The class Methanothermea needs careful testing; we do not know whether Methanopyrales and Methanomicrobiales have the split RNA polymerase B gene, as the other two orders do (Klenk & Zillig, 1994). The rRNA tree suggests that at least *Methanopyrus* might have branched before this innovation and therefore be better placed in Protoarchaea. If its most divergent position on rRNA trees is correct, then the ancestral euryarchaeote was probably a methanogen. Methanogenesis was almost certainly lost by the ancestor of Halobacteriales, but could have been lost several times. Even though the ancestral methanogen was probably, like the ancestral archaeobacterium, a hyperthermophile, at least one has secondarily gone to the opposite extreme and evolved adaptations for psychrophily (Lim *et al.*, 2000).

'Domain' is not a taxonomic category and should not be treated as one; it is a useful informal term, but the three domains have a very different status; on the present system, eubacteria are a paraphyletic grade but not a taxon, whereas Archaeobacteria (division or phylum) and Eukaryota (empire or superkingdom; Cavalier-Smith, 1998) are both holophyletic taxa but of very unequal rank. The term domain is convenient when, for simplicity, we wish to ignore these distinctions, but in some contexts they are important and the classical ranks more informative. Like clades and grades, the domain terminology is best treated as complementary to the classical Linnaean hierarchy of categories and not as a replacement or addition to it. Clades, taxa and grades serve different purposes in biology, as do classification and cladification (Cavalier-Smith, 1998).

The widespread practice of treating every highly divergent bacterial group on the rRNA tree as a separate phylum (division), or worse still 'kingdom', is unsound. Calling them candidate divisions (e.g. Hugenholtz *et al.*, 1998a; Dojka *et al.*, 2000) is somewhat better. But it is still basically unsound, because the words division, phylum or kingdom imply a rank; 'candidate group' would be better usage. One cannot sensibly judge the distinctiveness of a group or how it should be ranked solely on its depth of

branching or grouping in a single-gene tree. rRNA, though technically useful and properly much studied, is one of the worst molecules for making such judgements because it is much more subject to base-composition biases, length variations and interlineage rate variations than most proteins commonly used for megaphylogeny. As its evolutionary rate can vary over a thousandfold, ranking by its degree of dissimilarity would be absurd. I predict that most candidate 'divisions', when studied by good multiple-protein trees, will be found to belong in one or other of the phyla accepted here. However, I agree with Hugenholtz *et al.* (1998b) that it is too early to predict how many new phyla will be needed once we know much more about the great richness of uncultured lineages.

It is most unwise to base too much on a single molecule; we must get away from the attitude that tempts a highly distinguished scientist to say 'this rRNA tree is surely the most important single guide we will ever have to understanding genealogical relationships between organisms' (Doolittle, 1999b). Because I accept that rRNA trees are important, I have spent much of the past decade using them to help unravel eukaryote phylogeny. But it is unfortunately true that rRNA trees have also been the single most misleading and dogmatically misinterpreted source of evidence on such matters; like Janus, they have two faces – a benign and a malevolent one. Excessive belief in the rRNA tree led Schutz *et al.* (2000) to suggest that the cytochrome *bc* complex was laterally transferred from ϵ -Proteobacteria to *Aquifex*. But lateral transfer of such a macromolecular complex is inherently unlikely; they should have realized that the cytological evidence discussed above and the RNA polymerase tree (Klenk *et al.*, 1999) suggest strongly that it is the rRNA tree that misplaces *Aquifex*, while the cytochrome trees more accurately reflect organismal evolution. If that is so, then the cytochrome *b* and *c* trees appear to have no lateral transfers, just vertical descent.

rRNA enthusiasts also have a sad record of over-confidently denying the monophyly of groups that are robustly monophyletic by classical morphological criteria and prematurely asserting the early divergence of groups differing greatly in rRNA sequence (e.g. for all the orange rogue lineages in Fig. 6, plus numerous others). Thus Stackebrandt & Woese (1980) denied the monophyly of Spirochaetes and Field *et al.* (1988) that of the animal kingdom. The fact that complete rRNA sequences later supported their monophyly does not alter the fact that the morphological evidence against the premature conclusion was unwisely discounted. Other cases where morphology pointed to monophyly, but was more controversial, but where it was claimed to have been disproved by rRNA are the phylum Mycetozoa (apparently contradicted by numerous trees) and the kingdoms Plantae (contradicted by Bhattacharya *et al.*, 1990) and Chromista (contradicted by Bhattacharya *et al.*, 1991; Bhattacharya & Medlin, 1995; Oliveira & Bhattacharya, 2000).

Protein trees have clearly established the monophyly of Mycetozoa (Baldauf & Doolittle, 1997; Baldauf *et al.*, 2000) and Plantae *sensu* Cavalier-Smith 1981 (Moreira *et al.*, 2000; Baldauf *et al.*, 2000). The recent evidence from the duplication and retargeting of glyceraldehyde-phosphate dehydrogenase (Fast *et al.*, 2001) decisively shows chromalveolates (Chromista plus alveolates; Cavalier-Smith, 1999, 2000a, b) to be monophyletic (McFadden, 2000), making it highly probable that Chromista are also. When there is a major sudden radiation, as in the radiation of each of the kingdoms Plantae and Chromista into three major lineages closely following the origin of chloroplasts or the lateral transfer of a red-algal chloroplast, respectively, it is very difficult for sequence trees, whether rRNA or protein, to resolve their branching order. If one radiation closely follows another, as in that instance, most will simply mix the lineages, either at random or according to misleading systematic biases. Failure of a sequence tree to resolve a massive radiation is expected and common and should not be used, as it often is, to devalue other evidence that allows one to group the taxa by sensible criteria.

Grouping rRNA sequences by similarity is very useful in a database like GenBank when there is little or no other evidence of their true relationships, but such grouping of sequences should not be confused, as it often is, with a classification of the organisms. We know for several eukaryote groups (e.g. Mycetozoa) that the rRNA classification used by GenBank is wrong. This is also bound to be true for some bacterial groups. Each lineage that does not group obviously with one of the eight phyla accepted here should be examined intensively and critically by a spectrum of methods, morphological and biochemical, to determine whether, despite appearances from its rRNA sequence, it really belongs in an established phylum, or is a proper basis for a new phylum. The fact that most lineages fall within the phyla defined in terms of distinctive cell envelope structure and chemistry and photosynthetic and motility mechanisms means that these classical criteria are a very good basis for defining phyla (Cavalier-Smith, 1998) and that real bacterial organisms exist; they are not just a random assemblage of genes. The number of additional phyla that will be needed may be rather small.

Envoi

As there are 20 independent arguments that polarize the tree from eubacteria to neomura, the derived nature of neomura compared with eubacteria is no longer in doubt. The fossil evidence indicates strongly that neomura are only about a quarter as old as eubacteria. The idea that archaeobacteria and eukaryotes are both ancient (Woese & Fox, 1977; Woese, 1987, 2000) is firmly contradicted by all palaeontological and cell-biological evidence and is not required by any molecular evidence, so must be abandoned. So also must another serious misinterpretation of the universal tree of life: the idea that the

cenancestor was an ill-developed progenote (Woese & Fox, 1977; Woese, 1998, 2000); the evidence is compelling that it was a highly developed eubacterium and rather strong that it was a negibacterium. That it was specifically a green non-sulphur bacterium is the best current working hypothesis. Eobacteria, especially Chlorobacteria, which are unexpectedly diverse and widespread (Hugenholtz *et al.*, 1998b), need to be studied intensively and extensively to ascertain whether they really are the phylum that is most divergent from all other organisms, not just another case of our being fooled by great divergence on rRNA trees and of our confusing primitive absence of characters (lipopolysaccharide, flagella) with their secondary loss.

Our understanding of the origin of neomura would be enhanced by similarly extensive study of Actinobacteria, especially the classes Arabobacteria and Streptomycetes, the leading candidates for the ancestors of archaebacteria and eukaryotes. The fossil evidence must be given much more prominence in discussing the timing of evolutionary events; it places the reality of quantum evolution and mosaic evolution beyond question and falsifies the sometimes heuristically useful idea of the molecular clock. Many current interpretations in molecular evolution need to be re-evaluated carefully in the light of the extreme distortion of molecular trees that quantum evolution can cause and the palaeontologically sounder rerooting of the universal tree advocated here. I invite the strongest possible reasoned criticisms of this synthesis.

ACKNOWLEDGEMENTS

I thank NERC for a Professorial Fellowship and research grant, P.J. Keeling for suggesting that I look into protein synthesis initiation factor evolution, A.J. Roger for stimulating discussions and many valuable comments on the manuscript, numerous other members of the Evolutionary Biology Programme of the Canadian Institute for Advanced Research (CIAR) for useful discussions and CIAR itself for support as a Fellow.

REFERENCES

- Achenbach-Richter, L., Gupta, R., Stetter, K. & Woese, C. R. (1987). Were the original eubacteria thermophiles? *Syst Appl Microbiol* **9**, 34–39.
- Albers, S.-V., van de Vossenberg, J. L. C. M., Driessen, A. J. M. & Konings, W. N. (2000). Adaptations of the archaeal cell membrane to heat stress. *Frontiers Biosci* **5**, 796–803.
- Aravind, L. & Koonin, E. V. (2001). Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* **11**, 1365–1374.
- Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R. & Koonin, E. V. (1998). Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet* **14**, 442–444.
- Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R. & Koonin, E. V. (1999). Reply. *Trends Genet* **15**, 299–300.
- Aravind, L., Makarova, K. S. & Koonin, E. V. (2000). Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res* **28**, 3417–3432.
- Archibald, J. M., Logsdon, J. M. & Doolittle, W. F. (1999). Recurrent paralogy in the evolution of archaeal chaperonins. *Curr Biol* **9**, 1053–1056.
- Archibald, J. M., Logsdon, J. M., Jr & Doolittle, W. F. (2000). Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes. *Mol Biol Evol* **17**, 1456–1466.
- Archibald, J. M., Cavalier-Smith, T., Maier, U. & Douglas, S. (2001). Molecular chaperones encoded by a reduced nucleus: the cryptomonad nucleomorph. *J Mol Evol* **52**, 490–501.
- Atkins, J. F. & Gesteland, R. F. (2000). The twenty-first amino acid. *Nature* **407**, 463–465.
- Av-Gay, Y. & Everett, M. (2000). The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. *Trends Microbiol* **8**, 238–244.
- Ayala, F. J. (1999). Molecular clock mirages. *Bioessays* **21**, 71–75.
- Baldauf, S. L. & Doolittle, W. F. (1997). Origin and evolution of the slime molds (Mycetozoa). *Proc Natl Acad Sci USA* **94**, 12007–12012.
- Baldauf, S. L., Palmer, J. D. & Doolittle, W. F. (1996). The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc Natl Acad Sci USA* **93**, 7749–7754.
- Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**, 972–977.
- Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci USA* **93**, 9188–9193.
- Barton, N. H. (2000). Genetic hitchhiking. *Philos Trans R Soc Lond B Biol Sci* **355**, 1553–1562.
- Beaton, M. J. & Cavalier-Smith, T. (1999). Eukaryotic non-coding DNA is functional: evidence from the differential scaling of cryptomonad genomes. *Philos Trans R Soc Lond B Biol Sci* **266**, 2053–2059.
- Belfort, M. & Weiner, A. (1997). Another bridge between kingdoms: tRNA splicing in archaea and eukaryotes. *Cell* **89**, 1003–1006.
- Bendich, A. J. & Drlica, K. (2000). Prokaryotic and eukaryotic chromosomes: what's the difference? *Bioessays* **22**, 481–486.
- Bernander, R. (2000). Chromosome replication, nucleoid segregation and cell division in archaea. *Trends Microbiol* **8**, 278–283.
- Besendahl, A., Qiu, Y. L., Lee, J., Palmer, J. D. & Bhattacharya, D. (2000). The cyanobacterial origin and vertical transmission of the plastid tRNA(Leu) group-I intron. *Curr Genet* **37**, 12–23.
- Bhattacharya, D. & Medlin, L. (1995). The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. *J Phycol* **31**, 489–498.
- Bhattacharya, D., Elwood, H. J., Goff, L. J. & Sogin, M. L. (1990). Phylogeny of *Gracilaria lemaneiformis* (Rhodophyta) based on sequence analysis of its small subunit ribosomal RNA coding region. *J Phycol* **26**, 181–186.
- Bhattacharya, D., Medlin, L., Wainright, P. O., Ariztia, E. V., Bibeau, C., Stickel, S. K. & Sogin, M. L. (1991). Algae containing

- chlorophylls a+c are paraphyletic: molecular evolutionary analysis of the Chromophyta. *Evolution* **46**, 1801–1817.
- Blankenship, R. E. (1994).** Protein structure, electron transfer and evolution of prokaryotic photosynthetic reaction centers. *Antonie Leeuwenhoek* **65**, 311–329.
- Blankenship, R. E. (2001).** Molecular evidence for the evolution of photosynthesis. *Trends Plant Sci* **6**, 4–6.
- Blobel, G. (1980).** Intracellular protein topogenesis. *Proc Natl Acad Sci USA* **77**, 1496–1500.
- Boucher, Y. & Doolittle, W. F. (2000).** The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol Microbiol* **37**, 703–716.
- Brasier, M. D. (2000).** The Cambrian explosion and the slow burning fuse. *Sci Prog* **83**, 77–92.
- Brasier, M. D. & Lindsay, J. F. (1998).** A billion years of environmental stability and the emergence of eukaryotes: new data from northern Australia. *Geology* **26**, 555–558.
- Brasier, M., Green, O. & Shields, G. (1997).** Ediacarian sponge spicule clusters from southwestern Mongolia and the origins of the Cambrian fauna. *Geology* **25**, 303–306.
- Brinkmann, H. & Philippe, H. (1999).** Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Mol Biol Evol* **16**, 817–825.
- Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. (1999).** Archean molecular fossils and the early rise of eukaryotes. *Science* **285**, 1033–1036.
- Brown, J. R. & Doolittle, W. F. (1997).** Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev* **61**, 456–502.
- Brown, J. R. & Doolittle, W. F. (1999).** Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. *J Mol Evol* **49**, 485–495.
- Bruck, I. & O'Donnell, M. (2001).** The ring-type polymerase sliding clamp family. *Genome Biol* **2**, REVIEWS3001. <http://genomebiology.com/2001/2/1/reviews/3001/>
- Burggraf, S., Larsen, N., Woese, C. R. & Stetter, K. O. (1993).** An intron within the 16S ribosomal RNA gene of the archaeon *Pyrobaculum aerophilum*. *Proc Natl Acad Sci USA* **90**, 2547–2550.
- Burrows, J. A. & Goward, C. R. (1992).** Purification and properties of DNA polymerase from *Bacillus caldotenax*. *Biochem J* **287**, 971–977.
- Butterfield, N. J. (2000).** *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26**, 386–404.
- Butterfield, N. J., Knoll, A. H. & Swett, K. (1990).** A bangiophyte red alga from the Proterozoic of arctic Canada. *Science* **250**, 104–107.
- Cambillau, C. & Claverie, J. M. (2000).** Structural and genomic correlates of hyperthermostability. *J Biol Chem* **275**, 32383–32386.
- Canfield, D. E., Habicht, K. S. & Thamdrup, B. (2000).** The Archaeal sulfur cycle and the early history of atmospheric oxygen. *Science* **288**, 658–661.
- Cann, I. K., Ishino, S., Hayashi, I., Komori, K., Toh, H., Morikawa, K. & Ishino, Y. (1999).** Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. *J Bacteriol* **181**, 6591–6599.
- Castresana, J. & Moreira, D. (1999).** Respiratory chains in the last common ancestor of living organisms. *J Mol Evol* **49**, 453–460.
- Cavalier-Smith, T. (1975).** The origin of nuclei and of eukaryote cells. *Nature* **256**, 463–468.
- Cavalier-Smith, T. (1977).** Darwinism yesterday and today. *New Humanist* **92**, 182–185.
- Cavalier-Smith, T. (1978).** Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *J Cell Sci* **34**, 247–278.
- Cavalier-Smith, T. (1980).** Cell compartmentation and the origin of eukaryote membranous organelles. In *Endocytobiology: Endosymbiosis and Cell Biology, a Synthesis of Recent Research*, pp. 893–916. Edited by W. Schwemmler & H. E. A. Schenk. Berlin: De Gruyter.
- Cavalier-Smith, T. (1981).** The origin and early evolution of the eukaryotic cell. In *Molecular and Cellular Aspects of Microbial Evolution* (Society for General Microbiology Symposium no. 32), pp. 33–84. Edited by M. J. Carlile, J. F. Collins & B. E. B. Moseley. Cambridge: Cambridge University Press.
- Cavalier-Smith, T. (1983a).** Genetic symbionts and the origin of split genes and linear chromosomes. In *Endocytobiology II*, pp. 29–45. Edited by W. Schwemmler & H. E. A. Schenk. Berlin: De Gruyter.
- Cavalier-Smith, T. (1983b).** A 6-kingdom classification and a unified phylogeny. In *Endocytobiology II*, pp. 1027–1034. Edited by W. Schwemmler & H. E. A. Schenk. Berlin: de Gruyter.
- Cavalier-Smith, T. (1985a).** Introduction: the evolutionary significance of genome size. In *The Evolution of Genome Size*, pp. 1–6. Edited by T. Cavalier-Smith. Chichester: Wiley.
- Cavalier-Smith, T. (1985b).** DNA replication and the evolution of genome size. In *The Evolution of Genome Size*, pp. 211–251. Edited by T. Cavalier-Smith. Chichester: Wiley.
- Cavalier-Smith, T. (1985c).** Selfish DNA and the origin of introns. *Nature* **315**, 283–284.
- Cavalier-Smith, T. (1986).** The kingdoms of organisms. *Nature* **324**, 416–417.
- Cavalier-Smith, T. (1987a).** The origin of cells: a symbiosis between genes, catalysts, and membranes. *Cold Spring Harb Symp Quant Biol* **52**, 805–824.
- Cavalier-Smith, T. (1987b).** The origin of eukaryotic and archaeobacterial cells. *Ann NY Acad Sci* **503**, 17–54.
- Cavalier-Smith, T. (1987c).** The origin of Fungi and pseudofungi. In *Evolutionary Biology of the Fungi*, pp. 339–353. Symposium of the British Mycological Society, no. 13. Edited by A. D. M. Rayner, C. M. Brasier & D. Moore. Cambridge: Cambridge University Press.
- Cavalier-Smith, T. (1990).** Microorganism megaevolution: integrating the living and fossil evidence. *Rev Micropaleontol* **33**, 145–154.
- Cavalier-Smith, T. (1991a).** The evolution of cells. In *Evolution of Life*, pp. 271–304. Edited by S. Osawa & T. Honjo. Tokyo: Springer.
- Cavalier-Smith, T. (1991b).** The evolution of prokaryotic and eukaryotic cells. In *Fundamentals of Medical Cell Biology*, vol. 1, pp. 217–272. Edited by G. E. Bittar. Greenwich, CT: JAI Press.
- Cavalier-Smith, T. (1991c).** Intron phylogeny: a new hypothesis. *Trends Genet* **7**, 145–148.
- Cavalier-Smith, T. (1992a).** Origins of secondary metabolism. In *Secondary Metabolites: their Function and Evolution*, pp. 64–87. CIBA Foundation Symposium no. 171. Edited by D. J. Chadwick & J. Whelan. Chichester: Wiley.

- Cavalier-Smith, T. (1992b). Bacteria and eukaryotes. *Nature* **356**, 570.
- Cavalier-Smith, T. (1992c). Origin of the cytoskeleton. In *The Origin and Evolution of the Cell*, pp. 79–106. Edited by H. Hartman & K. Matsuno. Singapore: World Scientific Publishers.
- Cavalier-Smith, T. (1993). Evolution of the eukaryotic genome. In *The Eukaryotic Genome*, pp. 333–385. Edited by P. Broda, S. G. Oliver & P. Sims. Cambridge: Cambridge University Press.
- Cavalier-Smith, T. (1995). Membrane heredity, symbiogenesis, and the multiple origins of algae. In *Biodiversity and Evolution*, pp. 75–114. Edited by R. Arai, M. Kato & Y. Doi. Tokyo: National Science Museum Foundation.
- Cavalier-Smith, T. (1998). A revised six-kingdom system of life. *Biol Rev Camb Philos Soc* **73**, 203–266.
- Cavalier-Smith, T. (1999). Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol* **46**, 347–366.
- Cavalier-Smith, T. (2000a). Membrane heredity and early chloroplast evolution. *Trends Plant Sci* **5**, 174–182.
- Cavalier-Smith, T. (2000b). Flagellate megaevolution: the basis for eukaryote diversification. In *The Flagellates*, pp. 361–390. Edited by J. R. Green & B. C. Leadbeater. London: Taylor and Francis.
- Cavalier-Smith, T. (2000c). What are Fungi? In *The Mycota*, vol. VII, *Systematics and Evolution Part A*, pp. 3–37. Edited by D. J. McLaughlin, E. G. McLaughlin & P. A. Lemke. Berlin: Springer.
- Cavalier-Smith, T. (2001). Obcells as proto-organisms: membrane heredity, lithophosphorylation, and the origins of the genetic code, the first cells, and photosynthesis. *J Mol Evol* **53**, 555–595.
- Cavalier-Smith, T. (2002). The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Microbiol* (in press).
- Cavalier-Smith, T. & Beaton, M. J. (1999). The skeletal function of non-genic nuclear DNA: new evidence from ancient cell chimaeras. *Genetica* **106**, 3–13.
- Cavalier-Smith, T., Couch, J. A., Thorsteinsen, K. E., Gilson, P., Deane, J., Hill, D. A. & McFadden, G. I. (1996a). Cryptomonad nuclear and nucleomorph 18S rRNA phylogeny. *Eur J Phycol* **31**, 315–328.
- Cavalier-Smith, T., Allsopp, M. T. E. P., Chao, E. E., Boury-Esnault, N. & Vacelet, J. (1996b). Sponge phylogeny, animal monophyly and the origin of the nervous system: 18S rRNA evidence. *Can J Zool* **74**, 2031–2045.
- Chappe, B., Albrecht, P. & Michaelis, W. (1982). Polar lipids of archaeobacteria in sediments and petroleum. *Science* **217**, 65–66.
- Charette, M. & Gray, M. W. (2000). Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* **49**, 341–351.
- Chater, K. (1992). In *Secondary Metabolites: their Function and Evolution*, pp. 84. CIBA Foundation Symposium no. 171. Edited by D. J. Chadwick & J. Whelan. Chichester: Wiley.
- Chen, X., Quinn, A. M. & Wolin, S. L. (2000). Do ribonucleoproteins contribute to the resistance of *Deinococcus radiodurans* to ultraviolet irradiation. *Genes Dev* **14**, 777–782.
- Chistoserdova, L., Vorholt, J. A., Thauer, R. K. & Lidstrom, M. E. (1998). C₁ transfer enzymes and coenzymes linking methylo-
- trophic bacteria and methanogenic archaea. *Science* **281**, 99–102.
- Condo, I., Ciammaruconi, A., Benelli, D., Ruggero, D. & Londei, P. (1999). Cis-acting signals controlling translational initiation in the thermophilic archaeon *Sulfolobus solfataricus*. *Mol Microbiol* **34**, 377–384.
- Copeland, P. R., Fletcher, J. E., Carlson, B. A., Hatfield, D. L. & Driscoll, D. M. (2000). A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO J* **19**, 306–314.
- Counter, C. M., Meyerson, M., Eaton, E. N. & Weinberg, R. A. (1997). The catalytic subunit of yeast telomerase. *Proc Natl Acad Sci U S A* **94**, 9202–9207.
- Cousineau, B., Lawrence, S., Smith, D. & Belfort, M. (2000). Retrotransposition of a bacterial group II intron. *Nature* **404**, 1018–1021.
- Curnow, A. W., Hong, K. W., Yuan, R., Kim, S. I., Martins, O., Winkler, W., Henkin, T. M. & Soll, D. (1997). Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc Natl Acad Sci U S A* **94**, 11819–11826.
- Dawes, I. W. (1981). Sporulation in evolution. In *Molecular and Cellular Aspects of Microbial Evolution*, pp. 85–130. Edited by M. J. Carlile, J. F. Collins & B. E. B. Moseley. Cambridge: Cambridge University Press.
- De Beer, G. (1954). Archaeopteryx and evolution. *Adv Sci* **42**, 160–170.
- DeLong, E. F., King, L. L., Massana, R., Cittone, H., Murray, A., Schleper, C. & Wakeham, S. G. (1998). Dibiphytanyl ether lipids in nonthermophilic crenarchaeotes. *Appl Environ Microbiol* **64**, 1133–1138.
- Delwiche, C. F. & Palmer, J. D. (1996). Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Mol Biol Evol* **13**, 873–882.
- Dojka, M. A., Harris, J. K. & Pace, N. R. (2000). Expanding the known diversity and environmental distribution of an uncultured phylogenetic division of bacteria. *Appl Environ Microbiol* **66**, 1617–1621.
- Doolittle, R. F. (1995). The origins and evolution of eukaryotic proteins. *Philos Trans R Soc Lond B Biol Sci* **349**, 235–240.
- Doolittle, R. F. (1998). Microbial genomes opened up. *Nature* **392**, 339–342.
- Doolittle, R. F. & Handy, J. (1998). Evolutionary anomalies among the aminoacyl-tRNA synthetases. *Curr Opin Genet Dev* **8**, 630–636.
- Doolittle, W. F. (1978). Genes – in pieces: were they ever together? *Nature* **272**, 581–582.
- Doolittle, W. F. (1999a). Phylogenetic classification and the universal tree. *Science* **284**, 2124–2129.
- Doolittle, W. F. (1999b). Lateral genomics. *Trends Cell Biol* **9**, M5–M8.
- Doolittle, W. F. (2000). Uprooting the tree of life. *Sci Am* **282**, 90–95.
- Douglas, S. E., Murphy, C. A., Spencer, D. F. & Gray, M. W. (1991). Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* **350**, 148–151.
- Douglas, S., Zauner, S., Fraunholz, M. & 7 other authors (2001). The highly reduced genome of an enslaved algal nucleus. *Nature* **410**, 1091–1096.
- Dubnau, D. (1999). DNA uptake in bacteria. *Annu Rev Microbiol* **53**, 217–244.

- de Duve, C. (1996). The birth of complex cells. *Sci Am* **274**, 50–57.
- Edgell, D. R. & Doolittle, W. F. (1997). Archaea and the origin(s) of DNA replication proteins. *Cell* **89**, 995–998.
- Edgell, D. R., Malik, S. B. & Doolittle, W. F. (1998). Evidence of independent gene duplications during the evolution of archaeal and eukaryotic family B DNA polymerases. *Mol Biol Evol* **15**, 1207–1217.
- Edward, D. G. & Freundt, E. A. (1967). Proposal for *Mollicutes* as name of the class established for the order *Mycoplasmatales*. *Int J Syst Bacteriol* **17**, 267–268.
- Embley, T. M. & Hirt, R. P. (1998). Early branching eukaryotes? *Curr Opin Genet Dev* **8**, 624–629.
- Fagegaltier, D., Hubert, N., Carbon, P. & Krol, A. (2000). The selenocysteine insertion sequence binding protein SBP is different from the Y-box protein dbpB. *Biochimie* **82**, 117–122.
- Faguy, D. M. & Doolittle, W. F. (1998). Cytoskeletal proteins: the evolution of cell division. *Curr Biol* **8**, R338–R341.
- Faguy, D. M. & Doolittle, W. F. (2000). Horizontal transfer of catalase-peroxidase genes between archaea and pathogenic bacteria. *Trends Genet* **16**, 196–197.
- Faguy, D. M., Jarrell, K. F., Kuzio, J. & Kalmokoff, M. L. (1994). Molecular analysis of archaeal flagellins: similarity to the type IV pilin-transport superfamily widespread in bacteria. *Can J Microbiol* **40**, 67–71.
- Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. (2001). Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* **18**, 418–426.
- Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* **27**, 401–410.
- Feng, D. F. & Doolittle, R. F. (1997). Converting amino acid alignment scores into measures of evolutionary time: a simulation study of various relationships. *J Mol Evol* **44**, 361–370.
- Feng, D. F., Cho, G. & Doolittle, R. F. (1997). Determining divergence times with a protein clock: update and reevaluation. *Proc Natl Acad Sci USA* **94**, 13028–13033.
- Field, K. G., Olsen, G. J., Lane, D. J., Giovannoni, S. J., Ghiselin, M. T., Raff, E. C., Pace, N. R. & Raff, R. A. (1988). Molecular phylogeny of the animal kingdom. *Science* **239**, 748–753.
- Fitch, W. M. & Markowitz, E. (1970). An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochem Genet* **4**, 579–593.
- Forterre, P. (1995). Thermoreduction, a hypothesis for the origin of prokaryotes. *C R Acad Sci III* **318**, 415–422.
- Forterre, P. (1996). A hot topic: the origin of hyperthermophiles. *Cell* **85**, 789–792.
- Forterre, P. & Philippe, H. (1999). Where is the root of the universal tree of life? *Bioessays* **21**, 871–879.
- Forterre, P., Benachenhou-Lafha, N. & Labedan, B. (1993). Universal tree of life. *Nature* **362**, 795.
- Forterre, P., Bouthier De La Tour, C., Philippe, H. & Duguët, M. (2000). Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from archaea to bacteria. *Trends Genet* **16**, 152–154.
- Galtier, N., Tourasse, N. & Gouy, M. (1999). A non-hyperthermophilic common ancestor to extant life forms. *Science* **283**, 220–221.
- Gibbons, N. E. & Murray, R. E. (editors) (1978). *Bergey's Manual of Determinative Bacteriology*, 9th edn. Baltimore: Williams & Wilkins.
- Gilbert, W. (1986). The RNA world. *Nature* **319**, 618.
- Gilson, P. R. & McFadden, G. I. (1996). The miniaturized nuclear genome of a eukaryotic endosymbiont contains genes that overlap, genes that are cotranscribed, and the smallest known spliceosomal introns. *Proc Natl Acad Sci USA* **93**, 7737–7742.
- Glansdorff, N. (2000). About the last common ancestor, the universal life-tree and lateral gene transfer: a reappraisal. *Mol Microbiol* **38**, 177–185.
- Gogarten, J. P. & Kibak, H. (1992). The bioenergetics of the last common ancestor and the origin of the eukaryotic endomembrane system. In *The Origin and Evolution of the Cell*, pp. 131–162. Edited by H. Hartman & K. Masuno. Singapore: World Scientific Publishers.
- Gogarten, J. P., Kibak, H., Dittrich, P. & 8 other authors (1989). Evolution of the vacuolar H⁺ATPase: implications for the origin of eukaryotes. *Proc Natl Acad Sci USA* **86**, 6661–6665.
- Graham, D. E., Overbeek, R., Olsen, G. J. & Woese, C. R. (2000). An archaeal genomic signature. *Proc Natl Acad Sci USA* **97**, 3304–3308.
- Granick, S. (1965). Evolution of heme and chlorophyll. In *Evolving Genes and Proteins*, pp. 67–88. Edited by V. Bryson & H. J. Vogel. New York: Academic Press.
- Green, B. R. (2001). Was “molecular opportunism” a factor in the evolution of different photosynthetic light-harvesting pigment systems? *Proc Natl Acad Sci USA* **98**, 2119–2121.
- Gribaldo, S. & Cammarano, P. (1998). The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. *J Mol Evol* **47**, 508–516.
- Gribaldo, S., Lumia, V., Creti, R., de Macario, E. C., Sanangelantoni, A. & Cammarano, P. (1999). Discontinuous occurrence of the hsp70 (dnaK) gene among Archaea and sequence features of HSP70 suggest a novel outlook on phylogenies inferred from this protein. *J Bacteriol* **181**, 434–443.
- Grill, S., Gualerzi, C. O., Londei, P. & Blasi, U. (2000). Selective stimulation of translation of leaderless mRNA by initiation factor 2: evolutionary implications for translation. *EMBO J* **19**, 4101–4110.
- Gupta, R. S. (1998a). Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* **62**, 1435–1491.
- Gupta, R. S. (1998b). Life's third domain (Archaea): an established fact or an endangered paradigm? *Theor Popul Biol* **54**, 91–104.
- Gupta, R. S. (2000). The natural evolutionary relationships among prokaryotes. *Crit Rev Microbiol* **26**, 111–131.
- Gupta, R. S., Mukhtar, T. & Singh, B. (1999). Evolutionary relationships among photosynthetic prokaryotes (*Halobacterium chlorum*, *Chloroflexus aurantiacus*, cyanobacteria, *Chlorobium tepidum* and proteobacteria): implications regarding the origin of photosynthesis. *Mol Microbiol* **32**, 893–906.
- Hahn, J. & Haug, P. (1986). Traces of archaeobacteria in ancient sediments. *Syst Appl Microbiol* **7**, 178–183.
- Han, T.-M. & Runnegar, B. (1992). Megascopic eukaryotic algae from the 2.1-billion-year-old Negaunee iron-formation, Michigan. *Science* **257**, 232–235.
- Handy, J. & Doolittle, R. F. (1999). An attempt to pinpoint the

- phylogenetic introduction of glutaminyl-tRNA synthetase among bacteria. *J Mol Evol* **49**, 709–715.
- Hedlund, B. P., Gosink, J. J. & Staley, J. T. (1997). *Verrucomicrobia* div. nov., a new division of the bacteria containing three new species of *Prostheco bacter*. *Antonie Leeuwenhoek* **72**, 29–38.
- Herschlag, D. (1998). Ribozyme crevices and catalysis. *Nature* **395**, 548–549.
- Hilario, E. & Gogarten, J. P. (1998). The prokaryote-to-eukaryote transition reflected in the evolution of the V/F/A-ATPase catalytic and proteolipid subunits. *J Mol Evol* **46**, 703–715.
- Hirt, R. P., Logsdon, J. M., Jr, Healy, B., Dorey, M. W., Doolittle, W. F. & Embley, T. M. (1999). Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc Natl Acad Sci U S A* **96**, 580–585.
- Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. (1998). A neoproterozoic snowball earth. *Science* **281**, 1342–1346.
- Hori, H. T., Itoh, T. & Osawa, S. (1982). The phylogenetic structure of the metabacteria. *Zentbl Bakteriell Mikrobiol Hyg C* **3**, 18–30.
- Horken, K. M. & Tabita, F. R. (1999). The “green” form I ribulose 1,5-bisphosphate carboxylase/oxygenase from the nonsulfur purple bacterium *Rhodobacter capsulatus*. *J Bacteriol* **181**, 3935–3941.
- Horwich, A. L. & Saibil, H. R. (1998). The thermosome: chaperonin with a built-in lid. *Nat Struct Biol* **5**, 333–336.
- Hugenholtz, P., Pitulle, C., Hershberger, K. L. & Pace, N. R. (1998a). Novel division level bacterial diversity in a Yellowstone hot spring. *J Bacteriol* **180**, 366–376.
- Hugenholtz, P., Goebel, B. M. & Pace, N. R. (1998b). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**, 4765–4774.
- Huynen, M., Snel, B. & Bork, P. (1999). Lateral gene transfer, genome surveys, and the phylogeny of prokaryotes. *Science* **286**, 1443.
- Hyde, W. T., Crowley, T. J., Baum, S. K. & Peltier, W. R. (2000). Neoproterozoic ‘snowball Earth’ simulations with a coupled climate/ice-sheet model. *Nature* **405**, 425–429.
- Ivanovsky, R. N., Fal, Y. I., Berg, I. A., Ugolkova, N. V., Krasilnikova, E. N., Keppen, O. I., Zakharchuc, L. M. & Zyakun, A. M. (1999). Evidence for the presence of the reductive pentose phosphate cycle in a filamentous anoxygenic photosynthetic bacterium, *Oscillochloris trichoides* strain DG-6. *Microbiology* **145**, 1743–1748.
- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. (1989). Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci U S A* **86**, 9355–9359.
- Jain, R., Rivera, M. C. & Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A* **96**, 3801–3806.
- Jefferies, R. S. (1979). The origin of chordates: a methodological essay. In *The Origin of Major Invertebrate Groups*, pp. 443–477. Edited by M. R. House. London: Academic Press.
- Kamiya, R., Hotani, H. & Asakura, S. (1982). Polymorphic transition in bacterial flagella. In *Prokaryotic and Eukaryotic Flagella*, pp. 53–76. Edited by W. B. Amos & J. G. Duckett. Cambridge: Cambridge University Press.
- Kampranis, S. C. & Maxwell, A. (1996). Conversion of DNA gyrase into a conventional type II topoisomerase. *Proc Natl Acad Sci U S A* **93**, 14416–14421.
- Kasinsky, H. E., Lewis, J. D., Dacks, J. B. & Ausio, J. (2001). Origin of H1 linker histones. *FASEB J* **15**, 34–42.
- Kasting, J. F., Holland, H. D. & Kump, L. R. (1992). Atmospheric evolution: the rise of oxygen. In *The Proterozoic Biosphere*, pp. 159–163. Edited by J. W. Schopf & C. Klein. Cambridge: Cambridge University Press.
- Keeling, P. J. & Doolittle, W. F. (1995). Archaea: narrowing the gap between prokaryotes and eukaryotes. *Proc Natl Acad Sci U S A* **92**, 5761–5764.
- Keeling, P. J. & McFadden, G. I. (1998). Origins of microsporidia. *Trends Microbiol* **6**, 19–23.
- Keeling, P. J., Fast, N. M. & McFadden, G. I. (1998). Evolutionary relationship between translation initiation factor eIF-2 gamma and selenocysteine-specific elongation factor SELB: change of function in translation factors. *J Mol Evol* **47**, 649–655.
- Keeling, P. J., Luker, M. A. & Palmer, J. D. (2000). Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi. *Mol Biol Evol* **17**, 23–31.
- Kim, K. K., Hung, L. W., Yokota, H., Kim, R. & Kim, S. H. (1998). Crystal structures of eukaryotic translation initiation factor 5A from *Methanococcus jannaschii* at 1.8 Å resolution. *Proc Natl Acad Sci U S A* **95**, 10419–10424.
- Kimura, M. (1963). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- King, J. L. & Jukes, T. H. (1969). Non-Darwinian evolution. *Science* **164**, 788–798.
- Kirschvink, J. L., Gaidos, E. J., Bertani, L. E., Beukes, N. J., Gutzmer, J., Maepa, L. N. & Steinberger, R. E. (2000). Paleoproterozoic snowball earth: extreme climatic and geochemical global change and its biological consequences. *Proc Natl Acad Sci U S A* **97**, 1400–1405.
- Klenk, H. P. & Zillig, W. (1994). DNA-dependent RNA polymerase subunit B as a tool for phylogenetic reconstructions: branching topology of the archaeal domain. *J Mol Evol* **38**, 420–432.
- Klenk, H.-P., Clayton, R. A., Tomb, J.-F. & 48 other authors (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**, 364–370.
- Klenk, H. P., Meier, T. D., Durovic, P., Schwass, V., Lottspeich, F., Dennis, P. P. & Zillig, W. (1999). RNA polymerase of *Aquifex pyrophilus*: implications for the evolution of the bacterial rpoBC operon and extremely thermophilic bacteria. *J Mol Evol* **48**, 528–541.
- Knoll, A. H. (1992). The early evolution of eukaryotes: a geological perspective. *Science* **256**, 622–627.
- Kohl, W., Gloe, A. & Reichenbach, H. (1983). Steroids from the myxobacterium *Nannocystis exedens*. *J Gen Microbiol* **129**, 1629–1635.
- Kollman, J. M. & Doolittle, R. F. (2000). Determining the relative rates of change for prokaryotic and eukaryotic proteins with anciently duplicated paralogs. *J Mol Evol* **51**, 173–181.
- Kong, X. P., Onrust, R., O'Donnell, M. & Kuriyan, J. (1992). Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell* **69**, 425–437.
- Koonin, E. V., Mushegian, A. R., Galperin, M. Y. & Walker, D. R. (1997). Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Mol Microbiol* **25**, 619–637.

- Koonin, E. V., Wolf, Y. I. & Aravind, L. (2001). Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res* **11**, 240–252.
- Kowalak, J. A., Dalluge, J. J., McCloskey, J. A. & Stetter, K. O. (1994). The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles. *Biochemistry* **33**, 7869–7876.
- Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M. & Kuriyan, J. (1994). Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* **79**, 1233–1243.
- Kyrpides, N. C. & Olsen, G. J. (1999). Archaeal and bacterial hyperthermophiles: horizontal gene exchange or common ancestry? *Trends Genet* **15**, 298–299.
- Kyrpides, N. C. & Woese, C. R. (1998a). Universally conserved translation initiation factors. *Proc Natl Acad Sci USA* **95**, 224–228.
- Kyrpides, N. C. & Woese, C. R. (1998b). Archaeal translation initiation revisited: the initiation factor 2 and eukaryotic initiation factor 2B alpha-beta-delta subunit families. *Proc Natl Acad Sci USA* **95**, 3726–3730.
- Labadan, B., Boyen, A., Baetens, M. & 16 other authors (1999). The evolutionary history of carbamoyltransferases: a complex set of paralogous genes was already present in the last universal common ancestor. *J Mol Evol* **49**, 461–473.
- Lake, J. A. (1988). Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* **331**, 184–186.
- Lamb, D. C., Kelly, D. E., Manning, N. J. & Kelly, S. L. (1998). A sterol biosynthetic pathway in *Mycobacterium*. *FEBS Lett* **437**, 142–144.
- Lamour, V., Quevillon, S., Diriong, S., N'Guyen, V. C., Lipinski, M. & Mirande, M. (1994). Evolution of the Glx-tRNA synthetase family: the glutamyl enzyme as a case of horizontal gene transfer. *Proc Natl Acad Sci USA* **91**, 8670–8674.
- Landthaler, M. & Shub, D. A. (1999). Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. *Proc Natl Acad Sci USA* **96**, 7005–7010.
- Lechner, J., Wieland, F. & Sumper, M. (1986). Sulfated dolicholphosphate oligosaccharides are transiently methylated during biosynthesis of halobacterial glycoproteins. *Syst Appl Microbiol* **7**, 286–292.
- Lee, M. S. Y. (1999). Molecular clock calibrations and metazoan divergence dates. *J Mol Evol* **49**, 385–391.
- Lee, J. H., Choi, S. K., Roll-Mecak, A., Burley, S. K. & Dever, T. E. (1999). Universal conservation in translation initiation revealed by human and archaeal homologs of bacterial translation initiation factor IF2. *Proc Natl Acad Sci USA* **96**, 4342–4347.
- Leipe, D. D., Aravind, L. & Koonin, E. V. (1999). Did DNA replication evolve twice independently? *Nucleic Acids Res* **27**, 3389–3401.
- Leroux, M. R. & Hartl, F. U. (2000). Protein folding: versatility of the cytosolic chaperonin TRiC/CCT. *Curr Biol* **10**, R260–R264.
- Levy, M. & Miller, S. L. (1998). The stability of the RNA bases: implications for the origin of life. *Proc Natl Acad Sci USA* **95**, 7933–7938.
- Lewin, R. A. & Cheng, L. (1989). *Prochloron: a Microbial Enigma*. New York: Chapman & Hall.
- Li, C., Motaleb, A., Sal, M., Goldstein, S. F. & Charon, N. W. (2000). Spirochete periplasmic flagella and motility. *J Mol Microbiol Biotechnol* **2**, 345–354.
- Lim, J., Thomas, T. & Cavicchioli, R. (2000). Low temperature regulated DEAD-box RNA helicase from the Antarctic archaeon, *Methanococcoides burtonii*. *J Mol Biol* **297**, 553–567.
- Llorca, O., McCormack, E. A., Hynes, G., Grantham, J., Cordell, J., Carrascosa, J. L., Willison, K. R., Fernandez, J. J. & Valpuesta, J. M. (1999a). Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. *Nature* **402**, 693–696.
- Llorca, O., Smyth, M. G., Carrascosa, J. L., Willison, K. R., Radermacher, M., Steinbacher, S. & Valpuesta, J. M. (1999b). 3D reconstruction of the ATP-bound form of CCT reveals the asymmetric folding conformation of a type II chaperonin. *Nat Struct Biol* **6**, 639–642.
- Logsdon, J. M., Jr (1998). The recent origins of spliceosomal introns revisited. *Curr Opin Genet Dev* **8**, 637–648.
- Logsdon, J. M. & Faguy, D. M. (1999). *Thermotoga* heats up lateral gene transfer. *Curr Biol* **9**, R747–R751.
- Lopez, P., Forterre, P. & Philippe, H. (1999). The root of the tree of life in the light of the covarian model. *J Mol Evol* **49**, 496–508.
- López-García, P. (1999). DNA supercoiling and temperature adaptation: a clue to early diversification of life? *J Mol Evol* **49**, 439–452.
- Lovejoy, A. O. (1960). *The Great Chain of Being: a Study of the History of an Idea*. New York: Harper.
- Lwoff, A. (1944). *L'Evolution Physiologique: Étude des Pertes de Fonctions chez les Microorganismes*. Paris: Hermann et Cie.
- McFadden, G. I. (2000). Mergers and acquisitions: malaria and the great chloroplast heist. *Genome Biol* **1**, REVIEWS1026. <http://genomebiology.com/2000/1/4/reviews/1026/>
- McIlroy, D., Green, O. R. & Brasier, M. D. (1994). The world's oldest foraminiferans. *Microsc Anal* **147**, 13–15.
- Maier, U.-G., Douglas, S. & Cavalier-Smith, T. (2000). The nucleomorph genomes of cryptophytes and chlorarachniophytes. *Protist* **151**, 103–109.
- Margulis, L. (1970). *Origin of Eukaryotic Cells*. New Haven, CT: Yale University Press.
- Margulis, L. (1974). Five kingdom classification and the origin and evolution of cells. *Evol Biol* **7**, 45–78.
- Martin, W. (1999). Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *Bioessays* **21**, 99–104.
- Martin, W. & Müller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41.
- Mason, N., Ciufo, L. F. & Brown, J. D. (2000). Elongation arrest is a physiologically important function of signal recognition particle. *EMBO J* **19**, 4164–4174.
- Maupin-Furlow, J. A., Kaczowka, S. J., Ou, M. S. & Wilson, H. L. (2001). Archaeal proteasomes: proteolytic nanocompartments of the cell. *Adv Appl Microbiol* **50**, 279–338.
- Maynard Smith, J. & Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford: Oxford University Press.
- Mayr, E. (1998). Two empires or three? *Proc Natl Acad Sci USA* **95**, 9720–9723.
- Mayr, E. & Ashlock, P. D. (1969). *Principles of Systematic Zoology*, 2nd edn. New York: McGraw, Hill.
- van der Meer, M. T., Schouten, S., van Dongen, B. E., Rijpsstra, W. I., Fuchs, G., Damste, J. S., de Leeuw, J. W. & Ward, D. M.

- (2001). Biosynthetic controls on the ^{13}C contents of organic components in the photoautotrophic bacterium *Chloroflexus aurantiacus*. *J Biol Chem* **276**, 10971–10976.
- Mizutani, T. & Fujiwara, T. (2000). SBP, SECIS binding protein, binds to the RNA fragment upstream of the Sec UGA codon in glutathione peroxidase mRNA. *Mol Biol Rep* **27**, 99–105.
- Møller-Jensen, J., Jensen, R. B. & Gerdes, K. (2000). Plasmid and chromosome segregation in prokaryotes. *Trends Microbiol* **8**, 313–320.
- Moreira, D. & López-García, P. (1998). Symbiosis between methanogenic archaea and δ -proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J Mol Evol* **47**, 517–530.
- Moreira, D., Le Guyader, H. & Philippe, H. (2000). The origin of red algae and the evolution of chloroplasts. *Nature* **405**, 69–72.
- Mylykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H. & Forterre, P. (2000). Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* **288**, 2212–2215.
- Napoli, A., Kvaratskeli, M., White, M. F., Rossi, M. & Ciaramella, M. (2001). A novel member of the Bacterial–Archaeal regulator family is a nonspecific DNA-binding protein and induces positive supercoiling. *J Biol Chem* **276**, 10745–10752.
- Nelson, K. E., Clayton, R. A., Gill, S. R. & 26 other authors (1999). Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329.
- Nelson, K. E., Levy, M. & Miller, S. L. (2000). Peptide nucleic acids rather than RNA may have been the first genetic molecule. *Proc Natl Acad Sci USA* **97**, 3868–3871.
- Nesbø, C. L., L'Haridon, S., Stetter, K. O. & Doolittle, W. F. (2001). Phylogenetic analysis of two 'archaeal' genes in *Thermotoga maritima* reveal multiple transfers between archaea and bacteria. *Mol Biol Evol* **18**, 362–375.
- Nowitzki, U., Flechner, A., Kellermann, J., Hasegawa, M., Schnarrenberger, C. & Martin, W. (1998). Eubacterial origin of nuclear genes for chloroplast and cytosolic glucose-6-phosphate isomerase from spinach: sampling eubacterial gene diversity in eukaryotic chromosomes through symbiosis. *Gene* **214**, 205–213.
- Offner, S., Hofacker, A., Wanner, G. & Pfeifer, F. (2000). Eight of fourteen gvp genes are sufficient for formation of gas vesicles in halophilic archaea. *J Bacteriol* **182**, 4328–4336.
- Olendzenski, L., Liu, L., Zhaxybayeva, O., Murphey, R., Shin, D. G. & Gogarten, J. P. (2000). Horizontal transfer of archaeal genes into the Deinococcaceae: detection by molecular and computer-based approaches. *J Mol Evol* **51**, 587–599.
- Oliveira, M. C. & Bhattacharya, D. (2000). Phylogeny of the Bangiophycidae (Rhodophyta) and the secondary endosymbiotic origin of algal plastids. *Am J Bot* **87**, 482–492.
- Olson, J. M. & Pierson, B. K. (1987). Evolution of reaction centers in photosynthetic prokaryotes. *Int Rev Cytol* **108**, 209–248.
- Omer, A. D., Lowe, T. M., Russell, A. G., Ebhardt, H., Eddy, S. R. & Dennis, P. P. (2000). Homologs of small nucleolar RNAs in Archaea. *Science* **288**, 517–522.
- Orgel, L. E. (1998). The origin of life – a review of facts and speculations. *Trends Biochem Sci* **23**, 491–495.
- Ormerod, J. G., Kimble, L. K., Nesbakken, T., Torgersen, Y. A., Woese, C. R. & Madigan, M. T. (1996). *Heliobacterium fasciatum* gen. nov. sp. nov. and *Heliobacterium gestii* sp. nov.: endospore-forming heliobacteria from rice field soils. *Arch Microbiol* **165**, 226–234.
- Osada, Y., Saito, R. & Tomita, M. (1999). Analysis of base-pairing potentials between 16S rRNA and 5' UTR for translation initiation in various prokaryotes. *Bioinformatics* **15**, 578–581.
- Pace, N. R. (1991). Origin of life – facing up to the physical setting. *Cell* **65**, 531–533.
- Pace, N. R., Olsen, G. J. & Woese, C. R. (1986). Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* **45**, 325–326.
- Page, R. D. (2000). Extracting species trees from complex gene trees: reconciled trees and vertebrate phylogeny. *Mol Phylogenet Evol* **14**, 89–106.
- Palmer, J. D., Adams, K. L., Cho, Y., Parkinson, C. L., Qiu, Y. L. & Song, K. (2000). Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc Natl Acad Sci USA* **97**, 6960–6966.
- Paoli, G. C., Soyer, F., Shively, J. & Tabita, F. R. (1998). *Rhodobacter capsulatus* genes encoding form I ribulose-1,5-bisphosphate carboxylase/oxygenase (*cbbLS*) and neighbouring genes were acquired by a horizontal gene transfer. *Microbiology* **144**, 219–227.
- Paquin, B., Kathe, S. D., Nierzwicki-Bauer, S. A. & Shub, D. A. (1997). Origin and evolution of group I introns in cyanobacterial tRNA genes. *J Bacteriol* **179**, 6798–6806.
- Paquin, B., Heinfing, A. & Shub, D. A. (1999). Sporadic distribution of tRNA(Arg)CCU introns among alpha-purple bacteria: evidence for horizontal transmission and transposition of a group I intron. *J Bacteriol* **181**, 1049–1053.
- Pawlowski, J., Bolivar, I., Fahrni, J. F., de Vargas, C., Gouy, M. & Zaninetti, L. (1997). Extreme differences in rates of molecular evolution of foraminifera revealed by comparison of ribosomal DNA sequences and the fossil record. *Mol Biol Evol* **14**, 498–505.
- Philippe, H. & Adoutte, A. (1998). The molecular phylogeny of Eukaryota: solid facts and uncertainties. In *Evolutionary Relationships Among Protozoa*, pp. 25–56. Edited by G. H. Coombs, K. Vickerman, M. A. Sleight & A. Warren. London: Kluwer.
- Philippe, H. & Forterre, P. (1999). The rooting of the universal tree of life is not reliable. *J Mol Evol* **49**, 509–523.
- Plotz, B. M., Lindner, B., Stetter, K. O. & Holst, O. (2000). Characterization of a novel lipid A containing D-galacturonic acid that replaces phosphate residues. The structure of the lipid A of the lipopolysaccharide from the hyperthermophilic bacterium *Aquifex pyrophilus*. *J Biol Chem* **275**, 11222–11228.
- Poole, A., Jeffares, D. & Penny, D. (1999). Early evolution: prokaryotes, the new kids on the block. *Bioessays* **21**, 880–889.
- Poplawski, A., Grabowski, B., Long, S. E. & Kelman, Z. (2001). The zinc-finger domain of the archaeal MCM protein is required for helicase activity. *J Biol Chem Papers in Press*, published Oct 17 2001. DOI: 10.1074/jbc.M108519200.
- Porter, S. & Knoll, A. H. (2000). Testate amoebae in the Neoproterozoic era: evidence from vase-shaped microfossils in the Chuar group, Grand Canyon. *Paleobiology* **26**, 360–385.
- Preston, C. M., Wu, K. Y., Molinski, T. F. & DeLong, E. F. (1996). A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proc Natl Acad Sci USA* **93**, 6241–6246.
- Ranson, N. A., White, H. E. & Saibil, H. R. (1998). Chaperonins. *Biochem J* **333**, 233–242.
- Rasmussen, B. (2000). Filamentous microfossils in a 3,235-million-year-old volcanogenic massive sulphide deposit. *Nature* **405**, 676–679.

- Redfield, R. J. (1993).** Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation. *J Hered* **84**, 400–404.
- Reeve, J. N., Sandman, K. & Daniels, C. J. (1997).** Archaeal histones, nucleosomes, and transcription initiation. *Cell* **89**, 999–1002.
- Reysenbach, A.-L. & Cady, S. L. (2001).** Microbiology of ancient and modern hydrothermal systems. *Trends Microbiol* **9**, 79–86.
- Rivera, M. C. & Lake, J. A. (1992).** Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257**, 74–76.
- Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. (1998).** Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci USA* **95**, 6239–6244.
- Rizzotti, M. (2000).** *Early Evolution*. Basel: Birkhäuser.
- Robinson, H., Gao, Y. G., McCrary, B. S., Edmondson, S. P., Shriver, J. W. & Wang, A. H. (1998).** The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature* **392**, 202–205.
- Roger, A. J. (1999).** Reconstructing early events in eukaryotic evolution. *Am Nat* **154**, S146–S163.
- Roger, A. J., Keeling, P. J. & Doolittle, W. F. (1994).** Introns, the broken transposons. *Soc Gen Physiol Ser* **49**, 27–37.
- Rohmer, M., Bouvier, P. & Ourisson, G. (1980).** Non-specific lanosterol and hopanoid biosynthesis by a cell-free system from the bacterium *Methylococcus capsulatus*. *Eur J Biochem* **112**, 557–560.
- Rosing, M. T. (1999).** ¹³C-Depleted carbon microparticles in > 3700-Ma sea-floor sedimentary rocks from west Greenland. *Science* **283**, 674–676.
- Ruepp, A., Eckerskorn, C., Bogoy, M. & Baumeister, W. (1998).** Proteasome function is dispensable under normal but not under heat shock conditions in *Thermoplasma acidophilum*. *FEBS Lett* **425**, 87–90.
- Ruepp, A., Graml, W., Santos-Martinez, M. L. & 7 other authors (2000).** The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508–513.
- Saito, R. & Tomita, M. (1999).** Computer analyses of complete genomes suggest that some archaeobacteria employ both eukaryotic and eubacterial mechanisms in translation initiation. *Gene* **238**, 79–83.
- Samuelson, J. C., Chen, M., Jiang, F., Möller, I., Wiedmann, M., Kuhn, A., Phillips, G. J. & Dalbey, R. E. (2000).** YidC mediates membrane protein insertion in bacteria. *Nature* **406**, 637–641.
- Sandman, K. & Reeve, J. N. (1998).** Origin of the eukaryotic nucleus. *Science* **280**, 501–503.
- Sara, M. & Sleytr, U. B. (2000).** S-Layer proteins. *J Bacteriol* **182**, 859–868.
- Schidlowski, M. (2001).** Carbon isotopes as biogeochemical recorders of life over 3.8 Ga of earth history: evolution of a concept. *Precambrian Res* **106**, 117–134.
- Schopf, J. W. (1992).** Paleobiology of the Archaea. In *The Proterozoic Biosphere*, pp. 25–39. Edited by J. W. Schopf & C. Klein. Cambridge: Cambridge University Press.
- Schopf, J. W. (1993).** Microfossils of the Early Archaean Apex chert: new evidence of the antiquity of life. *Science* **260**, 640–646.
- Schopf, J. W. (1994).** Disparate rates, differing fates: tempo and mode of evolution changed from the Precambrian to the Phanerozoic. *Proc Natl Acad Sci USA* **91**, 6735–6742.
- Schopf, J. W. & Walter, M. R. (1983).** Archaean microfossils: new evidence of ancient microbes. In *Earth's Earliest Biosphere: its Origin and Evolution*, chapter 9, pp. 214–239. Edited by J. W. Schopf. Princeton: Princeton University Press.
- Schutz, M., Brugna, M., Lebrun, E. & 9 other authors (2000).** Early evolution of cytochrome bc complexes. *J Mol Biol* **300**, 663–675.
- Sedlmeier, R., Werner, T., Kieser, H. M., Hopwood, D. A. & Schmieger, H. (1994).** tRNA genes of *Streptomyces lividans*: new sequences and comparison of structure and organization with those of other bacteria. *J Bacteriol* **176**, 5550–5553.
- Shen, Y., Buick, R. & Canfield, D. E. (2001).** Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* **410**, 77–81.
- Siebert, R., Leroux, M. R., Scheufler, C., Hartl, F. U. & Moarefi, I. (2000).** Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. *Cell* **103**, 621–632.
- Simpson, G. G. (1944).** *Tempo and Mode in Evolution*. New York: Columbia University Press.
- Simpson, G. G. (1953).** *The Major Features of Evolution*. New York: Columbia University Press.
- Smith, C. M. & Steitz, J. A. (1997).** Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell* **89**, 669–672.
- Stackebrandt, E. & Woese, C. R. (1981).** The evolution of Prokaryotes. In *Molecular and Cellular Aspects of Microbial Evolution* (Society for General Microbiology Symposium no. 32), pp. 1–31. Edited by M. J. Carlile, J. F. Collins & B. E. B. Moseley. Cambridge: Cambridge University Press.
- Stackebrandt, E., Murray, R. G. E. & Trüper, H. G. (1988).** *Proteobacteria* classis nov., a name for the phylogenetic taxon that includes the “purple bacteria and their relatives”. *Int J Syst Bacteriol* **38**, 321–325.
- Stanier, R. Y. (1970).** Some aspects of the biology of cells and their possible evolutionary significance. In *Organization and Control in Prokaryotic and Eukaryotic Cells* (Society for General Microbiology Symposium no. 20), pp. 1–38. Edited by H. P. Charles & B. C. J. G. Knight. Cambridge: Cambridge University Press.
- Stanier, R. Y. (1974).** Division I. The Cyanobacteria. In *Bergey's Manual of Determinative Bacteriology*, 8th edn, p. 22. Edited by R. E. Buchanan & N. E. Gibbons. Baltimore: Williams & Wilkins.
- Stanier, R. Y. & Cohen-Bazire, G. (1977).** Phototrophic prokaryotes: the cyanobacteria. *Annu Rev Microbiol* **24**, 225–274.
- Stanier, R. Y. & Van Niel, C. B. (1962).** The concept of a bacterium. *Arch Mikrobiol* **42**, 17–35.
- Stiller, J. W. & Hall, B. D. (1999).** Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol Biol Evol* **16**, 1270–1279.
- Stiller, J. W., Duffield, E. C. & Hall, B. D. (1998).** Amitochondriate amoebae and the evolution of DNA-dependent RNA polymerase II. *Proc Natl Acad Sci USA* **95**, 11769–11774.
- Stoltzfus, A., Logsdon, J. M., Jr, Palmer, J. D. & Doolittle, W. F. (1997).** Intron ‘sliding’ and the diversity of intron positions. *Proc Natl Acad Sci USA* **94**, 10739–10744.
- Strauss, H., Des Marais, D. J., Hayes, J. M. & Summons, R. E. (1992).** The carbon-isotopic record. In *The Proterozoic Biosphere*, pp. 117–127. Edited by J. W. Schopf & C. Klein. Cambridge: Cambridge University Press.
- Stuart, R. A. & Neupert, W. (2000).** Making membranes in bacteria. *Nature* **406**, 575–577.

- Summons, R. E. & Hayes, J. M. (1992). Principles of molecular and isotopic biogeochemistry. In *The Proterozoic Biosphere*, pp. 83–93. Edited by J. W. Schopf & C. Klein. Cambridge: Cambridge University Press.
- Swan, D. G., Hale, R. S., Dhillon, N. & Leadlay, P. F. (1987). A bacterial calcium-binding protein homologous to calmodulin. *Nature* **329**, 84–85.
- Takaichi, S., Inoue, K., Akaike, M., Kobayashi, M., Oh-oka, H. & Madigan, M. T. (1997). The major carotenoid in all known species of heliobacteria is the C30 carotenoid 4,4'-diaponeurosporene, not neurosporene. *Arch Microbiol* **168**, 277–281.
- Teichmann, S. A. & Mitchison, G. (1999). Is there a phylogenetic signal in prokaryote proteins? *J Mol Evol* **49**, 98–107.
- Tenaillon, O., Toupance, B., Le Nagard, H., Taddei, F. & Godelle, B. (1999). Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics* **152**, 485–493.
- Thuret, G. (1875). Essai de classification des Nostochinées. *Ann Sci Nat Bot* **6**, 372–382.
- Tjalsma, H., Bolhuis, A., Jongbloed, J. D., Bron, S. & van Dijk, J. M. (2000). Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev* **64**, 515–547.
- Turner, S., Pryer, K. M., Miao, V. P. & Palmer, J. D. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J Eukaryot Microbiol* **46**, 327–338.
- Ugol'kova, N. V. & Ivanovskii, R. N. (2000). On the mechanism of autotrophic fixation of carbon dioxide by *Chloroflexus aurantiacus*. *Mikrobiologiya* **69**, 175–179 (in Russian).
- Van de Peer, Y., Ben Ali, A. & Meyer, A. (2000). Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. *Gene* **246**, 1–8.
- Van Valen, L. M. & Maiorana, V. C. (1980). The archaeobacteria and eukaryotic origins. *Nature* **287**, 248–250.
- Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vossbrinck, B. A. & Woese, C. R. (1987). Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature* **326**, 411–414.
- Walsh, M. M. & Lowe, D. R. (1985). Filamentous microfossils from the 3,500-M-yr-old Onverwacht Group, Barberton Mountain Land, South Africa. *Nature* **314**, 530–532.
- Walter, P., Keenan, R. & Schmitz, U. (2000). SRP – where the RNA and membrane worlds meet. *Science* **287**, 1212–1213.
- Ward, N. L., Rainey, F. A., Hedlund, B. P., Staley, J. T., Ludwig, W. & Stackebrandt, E. (2000). Comparative phylogenetic analyses of members of the order *Planctomycetales* and the division *Verrucomicrobia*: 23S rRNA gene sequence analysis supports the 16S rRNA gene sequence-derived phylogeny. *Int J Syst Evol Microbiol* **50**, 1965–1972.
- Watanabe, Y. & Gray, M. W. (2000). Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in *Euglena gracilis*, an early diverging eukaryote, and candidate Garlp and Nop10p homologs in archaeobacteria. *Nucleic Acids Res* **28**, 2342–2352.
- Westall, F., de Wit, M. J., Dann, J., van der Gaast, S., de Ronde, C. E. J. & Gerneke, D. (2001). Early Archaean fossil bacteria and biofilms in hydrothermally-influenced sediments from the Barberton greenstone belt, South Africa. *Precambrian Res* **106**, 93–116.
- Woese, C. R. (1982). Archaeobacteria and cellular origins: an overview. *Zentralbl Bakteriol Hyg 1 Abt Orig C* **3**, 1–17.
- Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev* **51**, 221–271.
- Woese, C. R. (1994). There must be a prokaryote somewhere: microbiology's search for itself. *Microbiol Rev* **58**, 1–9.
- Woese, C. R. (1998). The universal ancestor. *Proc Natl Acad Sci USA* **95**, 6854–6859.
- Woese, C. R. (2000). Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci USA* **97**, 8392–8396.
- Woese, C. R. & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* **74**, 5088–5090.
- Woese, C. R. & Gupta, R. (1981). Are archaeobacteria merely derived 'prokaryotes'? *Nature* **289**, 95–96.
- Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* **87**, 4576–4579.
- Woese, C. R., Olsen, G. J., Ibba, M. & Soll, D. (2000). Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev* **64**, 202–236.
- Wolf, Y. I., Aravind, L., Grishin, N. V. & Koonin, E. V. (1999). Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res* **9**, 689–710.
- Xi, C., Schoeters, E., Vanderleyden, J. & Michiels, J. (2000). Symbiosis-specific expression of *Rhizobium etli* *casA* encoding a secreted calmodulin-related protein. *Proc Natl Acad Sci USA* **97**, 11114–11119.
- Xiong, J., Inoue, K. & Bauer, C. E. (1998). Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from *Heliobacillus mobilis*. *Proc Natl Acad Sci USA* **95**, 14851–14856.
- Xiong, J., Fischer, W. M., Inoue, K., Nakahara, M. & Bauer, C. E. (2000). Molecular evidence for the early evolution of photosynthesis. *Science* **289**, 1724–1730.
- Xue, H., Guo, R., Wen, Y., Liu, D. & Huang, L. (2000). An abundant DNA binding protein from the hyperthermophilic archaeon *Sulfolobus shibatae* affects DNA supercoiling in a temperature-dependent fashion. *J Bacteriol* **182**, 3929–3933.
- Zauner, S., Fraunholz, M., Wastl, J., Penny, S., Beaton, M., Cavalier-Smith, T., Maier, U.-G. & Douglas, S. (2000). Chloroplast protein and centrosomal genes, a tRNA intron, and odd telomeres in an unusually compact eukaryotic genome, the cryptomonad nucleomorph. *Proc Natl Acad Sci USA* **97**, 200–205.
- Zhang, Z., Green, B. R. & Cavalier-Smith, T. (1999). Single gene circles in dinoflagellate chloroplast genomes. *Nature* **400**, 155–159.
- Zhang, Z., Green, B. R. & Cavalier-Smith, T. (2000). Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids. *J Mol Evol* **51**, 26–40.
- Zhu, B. C. & Laine, R. A. (1996). Dolichyl-phosphomannose synthase from the archae *Thermoplasma acidophilum*. *Glycobiology* **6**, 811–816.