The nature of the last universal common ancestor

David Penny* and Anthony Poole†

Cytologically, prokaryotes appear simpler and thus evolutionarily 'older' than eukaryotes. In terms of RNA processing, however, prokaryotes are sophisticated and eukaryotes, which retain many features of an RNA-world, appear primitive. The last universal common ancestor may have been mesophilic and could have had many features of the eukaryote genome, but its cytology is unknown.

Addresses
*Institute of Molecular BioSciences, Massey University, PO Box 11 222, Palmerston North, New Zealand; e-mail: D.Penny@massey.ac.nz
Department of Molecular Biology, Stockholm University, 106 91 Stockholm, Sweden; e-mail: anthony.poole@molbio.su.se

Current Opinion in Genetics & Development 1999, **9**:672–677

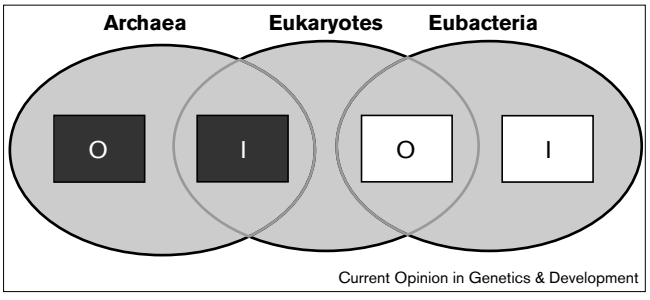
0959-437X/99/$ – see front matter

Abbreviations
LUCA last universal common ancestor
Mya Million years ago
RNP ribonucleoprotein

Introduction

The last universal common ancestor (LUCA) is the organism at the root of the 'tree of life' the ancestor of all organisms alive today [1••]. Estimating its properties helps understand even earlier steps in the origin of life — it is a window further back into time. Indeed, the usefulness of models of the LUCA are judged by their ability to improve understanding of these earlier steps, and models that make the origin of life more difficult to resolve require reconsideration.

Figure 1



Venn diagram of gene classes and their distribution. The broad consensus from whole-genome data is that the tree of life can best be understood by dividing all genes into two classes: informational (I) and operational (O). Although all organisms obviously require both classes, eukaryotes and archaea appear to resemble one another on the basis of similar informational genes, whereas eubacteria and eukaryotes group together at the exclusion of the eubacteria for operational genes. However, caution is required because 'similarity' is affected by differences in rates of evolution. Black and white boxes are included for clarity: black for I and O genes from archea, white for those from eubacteria.

The starting point for the tree of life has been the tripartite division into the domains archaea, bacteria and eukarya, initially on the basis of 16S rRNA sequences [2]. The horizontal transfer of genes between prokaryotes showed that the situation could be more complex [3••,4•,5•]. On the basis of function, genes are now classed as either informational or operational (Figure 1) [6••], Informational genes are broadly defined as those involved in informational processes, such as transcription, translation and replication, while operational genes are those that code for metabolic functions, such as enzymes involved in pathways for synthesis or breakdown of metabolites. The general consensus is that informational genes are less likely to be transferred because their products usually form large multicomponent complexes [3••], decreasing the likelihood of full interaction with several other macromolecules. There is considerable debate about the extent of horizontal (or lateral) transfer: some work favours virtually none [7••], whereas other authors favour huge transfers [8]. For the purpose of review, we accept the basic tripartite division, whilst still accepting significant horizontal transfer.

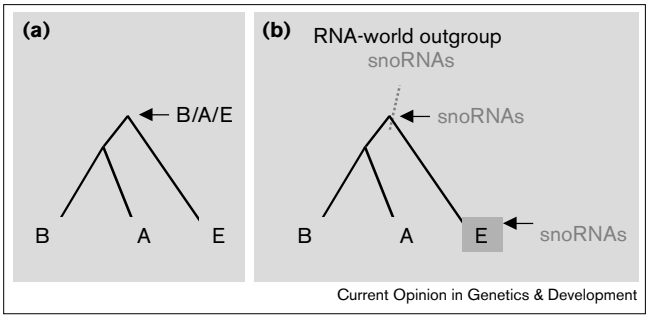
Assuming the tripartite division, the question of the nature of the LUCA is two-fold: identifying the position of the root of the tree of life and inferring its properties. The root has six possible placings, three on branches leading to each of the three domains, and three within any of the domains. We discuss two approaches to rooting the tree of life: analysis of sequence data (from single genes to complete genomes), and using inferred properties of the RNA-world as an outgroup. Finally, we emphasise the requirement for applying known evolutionary mechanisms and selective forces in building models of early evolution. This makes the conclusions more general and increases the power of the models.

Rooting the tree of life

Sequence data

The standard method for rooting the tree of life is to build from genes that duplicated before the LUCA arose; two separate elongation factors [9] and two ATPases [10], each found in all three domains, were initially used, both giving a tree where the root lies between the bacteria and archaea-eukarya, and therefore suggesting the LUCA was bacterial-like. They both gave a tree in favour of a bacterial LUCA, the root lying between the bacteria and archaea–eukarya. This theory has recently been challenged by Philippe and co-workers [11••,12••] using an improved method for building trees for ancient divergences. They built trees for each domain separately and from these subtrees ranked each site on its number of changes — from the slowest to the fastest evolving. The next step is to build the full tree for all three domains, starting first with just the very slowest sites. These sites located the root as occurring

Figure 2



RNA-world as an outgroup. If the tree of life is rooted between the eukaryotes and archaea–bacteria then the nature of the LUCA is not determined just from existing domains – it could be any number of features of the existing groups (as in **[a]**). If a feature is inferred to be in the RNA-world (e.g. snoRNAs), however, then it is an outgroup in that it helps determine the nature of the LUCA (as in **[b]**). A, Archaea; B, Bacteria; E, Eukarya.

between eukarya and archaea–bacteria. On the step-wise addition of faster-evolving sites, the root eventually reverts to between bacteria and archaea–eukarya. The interpretation is that the slowest evolving sites are giving the correct position, and that the fastest sites succumb to the well-known 'long branches attract' phenomenon, eventually overwhelming the slowest-evolving sites. (See [13] for a review of long branch attraction.)

These findings are challenging, though we caution against over-interpreting analyses that involve either a single or a small number of genes. It is well known that, even for mammals, which diverged over the last 100 million years, different genes give different (though statistically similar) trees [14••]. Given numerous other cases where gene trees are ambiguous [13], it is not to be expected that the relationships between all organisms on Earth (which diverged 3000–4000 million years ago [Mya]) can be determined unambiguously from small numbers of genes. Problematic genes could be removed from the analysis though a fundamental problem remains [15••]: any site that was free to evolve over the whole period at, say, 0.5% change per every million years will become saturated with 20–40 changes per site. Detecting phylogenetic signal above noise for deep divergences is thus difficult, making many proteins unsuitable for such phylogenetic studies [15••]. Other factors worsen recovery: rate differences between lineages; long branch attraction; horizontal transfer; unrecognised gene duplications; changes in nucleotide frequency; and changes in functional constraints [13,15••,16••].

Perhaps tree-building can be improved. Existing methods incorrectly assume that each site retains its characteristic rate of evolution across the whole tree but secondary and tertiary structure evolves through time under a covarion model [17•]. It has been proven for some proteins that the distribution of changes between sites requires them to vary in rate [17•] and although this is unexplored it may give more reliable trees than current theory predicts. Despite improved phylogenetic methods [11••,12••], alternative systems for locating the root are desirable. An alternative approach, examining molecular fossils from an earlier stage of life, is considered next. Molecular fossils, or 'relics' can be broadly defined as those parts of modern metabolism which have persisted from an earlier stage in evolution. An advantage over actual fossils is that these are 'living fossils', of clear relevance to modern biology, though a difficulty is that, since the time of their origins, they have been modified greatly or partially replaced, prompting analogies to a palimpsest and the lingering smile of the Cheshire Cat!

'Molecular fossils' from the RNA-world

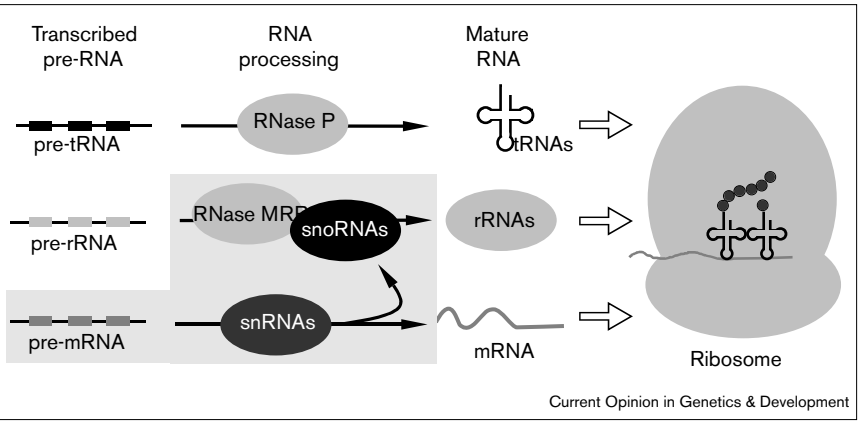
Knowledge of the properties of organisms that preceded the LUCA thus permits a different approach. The logic is straightforward (Figure 2) but the question is to identify such 'relics'. The furthest back we can extrapolate from modern life is to an RNA-world, where RNA carried out both coding and catalysis [18,19••,20••]. As they pre-date protein, any molecular fossil would be RNA-based and ancient in function. A modern RNA is deemed a possible relic if it is catalytic, or ubiquitous, or has a central role in metabolism, or carries out a role equally well served by a protein in other organisms [21••].

This approach is still relatively unexplored, though a fairly detailed model is possible [21••]. As depicted in Figure 3, the largest relic that can be reconstructed centres around the ribosome and translation. A general pattern exists where RNPs (ribonucleoproteins) process precursor RNAs, yielding mature RNA [19••]. In eukaryotes, this processing pattern is most complete, with tRNA, rRNA and mRNA each being produced via an RNP-processing pathway (Figure 3). Prokaryotes lack parts of this relic- processing pathway (shaded area in Figure 3) [19••]. Indeed, we consistently find all putative RNA-world relics in eukaryotes, with only a subset remaining in prokaryotes.

A separate issue is the application of information theory to understanding the origins of genome architecture in terms of replication fidelity. Lower fidelity favours multiple genome copies, a fragmented genome (as in chromosomes), and strong selection for recombination repair [22,23], and is indicative of the earliest replication systems. Eukaryotic genomes exhibit these hallmark low-fidelity traits — in contrast, prokaryotes usually house all information on one chromosome and are generally haploid, so additional copies cannot mask mutation. Nor can mutation be repaired by recombination in the absence of a homologous copy of the information.

It is difficult to consider prokaryote genome architecture and RNA processing as ancestral to that of eukaryotes as prokaryotic genome architecture is high fidelity with little RNA processing. This result and the new rooting the tree

Figure 3



The RNA processing pattern in eukaryotes reflects that of the LUCA. An examination of RNAs involved in translation reveals a striking pattern. Precursor RNAs are processed by RNPs (ribonucleoproteins–RNA plus cognate protein) to yield mature RNAs. Furthermore, RNPs process other RNPs – snoRNAs are released by sn RNAs, the RNA component of the splicing machinery, which in turn are crucial for rRNA processing. In prokaryotes, some of these RNAs have been lost (shaded region), and indeed, in the case of pre-mRNA, the processing step has been lost completely. Eukaryotes have retained a more complete record of the supposed RNA-world processing pathway than have prokaryotes.

of life between eukarya and archaea-bacteriais consistent with the conclusion that the genome architecture of the LUCA more closely resembled that of eukarya.

Thermoreduction and prokaryote origins

In postulating the nature of the LUCA, it is essential to consider the selective forces that would give rise to either prokaryotes or eukaryotes. Two selective forces that reinforce each other have been proposed by which prokaryotes could have evolved from an ancestor containing a eukaryote-like genome: thermoreduction and *r*-selection, [20••,21••,24]. *r*-selected organisms are fast-growing, competing for nutrient sources which fluctuate greatly in abundance. Yeast is *r*-selected when compared to an oak tree, which grows slowly, has a slow generation time and a fairly constant nutrient source (and is thus K-selected), and prokaryotes are *r*-selected relative to eukaryotes. *r* selection generally results in extremely fast and efficient use of resources, because limited availability produces strong competition for these. At the molecular level, the result is that enzymes that affect metabolite utilisation and organismal growth rate will be driven toward perfection at a faster rate than in organisms not under *r* selection. Thus, *r* selection may at least account partially for the observed replacement of RNA enzymes by protein in the prokaryote lineages [20••,21••].

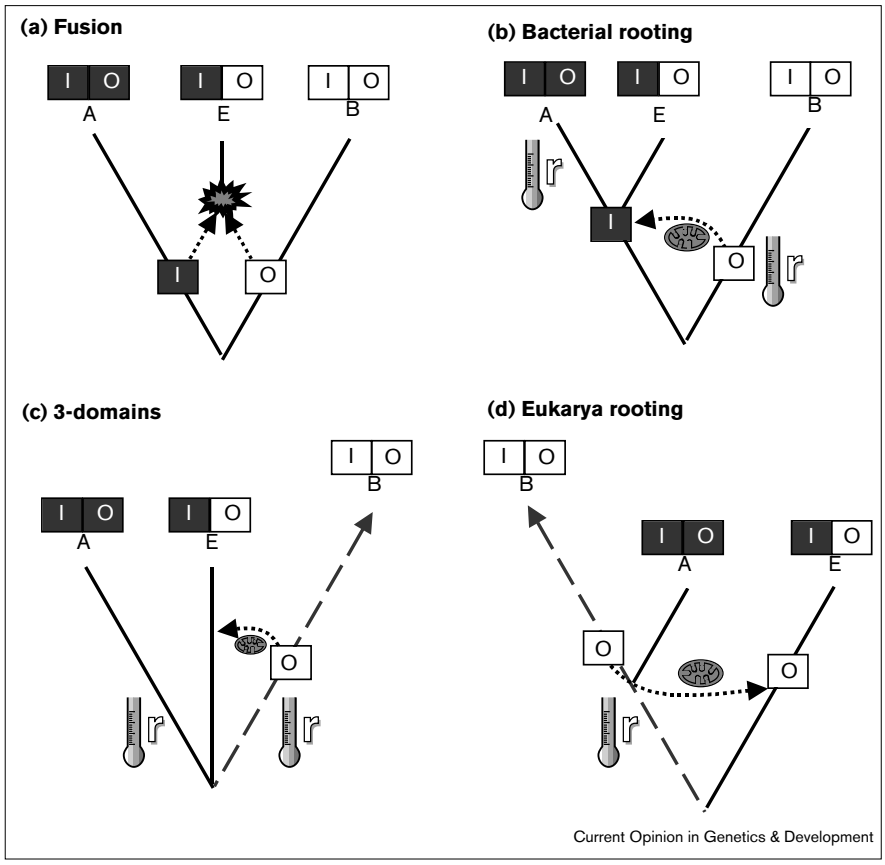
The thermoreduction hypothesis [24] is that prokaryotes arose from mesophiles by adaptation, via the loss of thermolabile traits, to high-temperature environments. This explains the loss of the ssRNA processing pathways (Figure 3) dating back to the RNA-world. Single-stranded RNA is heat labile, and would have been the Achilles' heel of early thermophiles. Accelerating ssRNA processing (mRNA, tRNA and rRNA) from hours (eukaryotes) to minutes (prokaryotes) would increase the viability of an organism at high temperatures. This loss of pre-mRNA processing, as well as the replacement of snoRNA-mediated rRNA processing with a protein enzyme system, would have been important steps in the evolution of thermophily. Unlike RNA, proteins are capable of extreme thermostability [25]. Furthermore, circular chromosomes are more thermostable than linear [26].

Other important molecules, such as glutamine [27] and carbamoyl phosphate [28], are also thermolabile. Glutamine is a protein amino acid and major nitrogen donor whereas carbamoyl phosphate is a crucial intermediate in the formation of pyrimidines and arginine. Pathways where carbamoyl phosphate and/or glutamine are used may have been affected by thermoreduction. For instance, in the hyperthermophilic archaeon *Pyrococcus furiosus*, carbamoyl phosphate is used immediately after synthesis by metabolite channelling, and has ammonia rather than glutamine as amino donor [28]. A second example of metabolite channelling is mischarging of glutaminyl–tRNA with glutamate, thereby making glutamine synthesis the final step before incorporation into protein; this is widespread within the prokaryotes but absent from eukaryotes [20••]. Although the area requires more investigation, the distribution of these traits in archaea and bacteria is predicted by the thermoreduction hypothesis.

Another dataset consistent with the LUCA being mesophilic comes from reconstructions of ancestral GC content. Galtier *et al.* [29••] have estimated its GC content and find it much lower than that characteristic of thermophiles. Moreover, a comparable result was obtained using only the thermophiles in their dataset. All work involving ancient sequence comparisons needs to be rigorously scrutinised but, in light of all the above data, the result is compelling nonetheless. In addition, that nucleotides themselves are unstable at high temperatures [30•] is consistent with a more mesophilic origin of life.

Overall, the thermoreduction hypothesis predicts a mesophilic LUCA with a genome and RNA-processing system more characteristic of eukarya. The power of the thermoreduction hypothesis is that it predicts a range of phenomena, rather than relying on *ad hoc* explanations of individual phenomena. Fossil dates do not contradict this picture because rocks from 2700

Figure 4



Fitting the data to the trees. Given our current understanding, several alternative trees could fit the data without altering either main conclusion. These are that the eukarya retain the greatest amount of biochemical similarity to the LUCA and that the prokaryotes have been through a period of reductive evolution, mainly through evolving to life at high temperatures. Some possible trees are as follows (episodes of thermoreduction and the origin of mitochondria are indicated). **(a)** The origin of eukarya (E) by fusion of a bacterium (B) and an archeon (A) fits the informational (I) and operational (O) gene distribution but is hard to fit all the data. It does not explain the origin of the nuclear membrane, however, which is assembled and disassembled during cell division, quite unlike organellar membranes (see [21••]). **(b)** Rooting the tree in the bacterial branch fits the data provided the biochemistry of the LUCA is understood to be more closely similar to that of modern eukaryotes than that of eubacteria. A bacterial rooting would require that the archaea and bacteria arose independently via r-selection and thermoreduction. **(c)** The classic 3-domain tree can also fit the data, provided the greater divergence of bacterial informational genes can be ascribed to higher rates of evolution. There would be transfer of operational genes back into eukarya through endosymbiosis. **(d)** The tree where the root is on the eukarya branch is perhaps the simplest with respect to the biochemical data. It is consistent with all the other data, provided (as for [c]) that the bacterial informational genes are indeed evolving at a faster rate.

Mya appear to have organic molecules characteristic of both prokaryotes and eukaryotes retained [31••].

Integrating data from genomes

Although data gleaned from biochemical approaches allows tentative reconstruction of the 'bare bones' LUCA, whole genomes will ultimately uncover much more information. Genomics allows metabolic traits to be compared through the presence or absence of genes, and by sequence comparisons. However, simple comparison of the presence or absence of homologous genes does not take into account the problems of gene loss or acquisition by horizontal transfer. Initial reconstruction of the 'minimal gene set' [32] highlights this caveat: being criticised because it resulted in exclusion of *de novo* pathways for deoxyribonucleotide synthesis, leading the authors to conclude that the LUCA had an RNA genome [33].

There is a difference between reconstructing the minimal gene set for cellular life, and the set of genes which the LUCA had. Greater caution is required when examining all three domains, as eukaryotes received prokaryotic genes subsequent to the endosymbioses of mitochondria and chloroplasts [34••,35••]. Replacement of unrelated, distantly related, or paralogous genes by functional counterparts is 'non-orthologous displacement' [36] and 'may' be central to understanding how the existing distribution of genes has arisen.

If we expect a eukaryote-like genome for LUCA as a starting point, how does this then fit with the data on operational and informational genes (Figure 1)? It is necessary to identify the direction of transfer. The complexity hypothesis [3••] places limits on gene transfer, such that we expect the transfer of mostly the operational genes in explaining the apparent chimerism. It has been suggested that acquisition of prokaryotic operational genes by eukaryotes results from their diet [37•]. There is no apparent selective advantage to such uptake, however, even though the mechanism might contribute to gene acquisition.

Another possibility is that the eukarya received the largest number of bacterial operational genes from the mitochondrion [38••]. Two established evolutionary mechanisms together favour this and are compatible with a eukaryal root: the increased rate of evolution toward catalytic perfection under r-selection [19••,21••], and Müller's ratchet.

Müller's ratchet is the term given to the continual accumulation of slightly deleterious mutations in lineages lacking recombination. It has been shown that Müller's ratchet is active in organelles [39] and that it drives the gene loss there (and also from obligate intracellular parasites) [34••,35••]. Most importantly, relocation of organellar genes to the nucleus benefits both host and symbiont. If the action of Müller's ratchet on the organelle drives gene loss, this can compromise the host-endosymbiont relationship and thus there is selection to relocate useful genes to the nucleus, where mutation rate is lower. The majority of endosymbiont genes were not expected to fit this category, however, and it was assumed these were lost over time, since equivalent functions already resided in the nucleus; but the simplest explanation of the evidence is that many were transferred [38••].

Figure 4 illustrates that the bioinformatic data, the RNA relic data, plus the evolutionary mechanisms that gave rise to the three domains can still fit several trees. Thus even with the nature of the LUCA, the branching order of the universal tree is not yet sufficiently informative to resolve all the issues. This is because each domain is a monophyletic group, so the basal branches of the tree (dividing the domains) can only take on a very limited number of trees. Hence, the metabolic data set cannot be used as an unambiguous outgroup for rooting the tree.

Conclusions

An interesting picture of the LUCA is emerging. It was a fully DNA and protein-based organism with extensive processing of RNA transcripts by RNPs (Figure 3). It had an extensive set of proteins for DNA, RNA and protein synthesis, DNA repair, recombination, control systems for regulation of genes and cell division, chaperone proteins, and probably lacked operons. Biochemistry favours a mesophilic LUCA with eukaryote-like RNA processing, though it is still possible to fit the data to several different trees (Figure 4). A eukaryote-like LUCA is not a new idea and can be traced back to Reanney [40].

Details of energy source(s) are unclear, partly because operational genes apparently undergo frequent horizontal transfers. Comparative genomics promises a clearer picture but apparent intermingling of lineages via horizontal transfer is a major obstacle [38••]. Increasingly, models need to fit our understanding of evolutionary theory and population genetics -it is essential to have plausible mechanisms and selective forces. The extent and direction of horizontal gene transfer needs accurate estimates before concluding the theory of descent does not hold for the earliest divergences [8,42,43]. Nevertheless, it is unclear whether the LUCA was a single 'species' or whether there was extensive horizontal transfer between divergent life forms. An outstanding issue is the origin of nuclear/cytoplasmic compartmentation as the concentration of RNA relics within the nucleus suggests this organelle is more ancient than previously supposed.

Acknowledgements

We thank H Philippe, P Forterre, H Brinkmann and P Lopez for sending manuscripts prior to publication.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Doolittle WF: **Phylogenetic classification and the universal tree.** •• *Science* 1999, **284**:2124-2128.

A recent review of developments in the fields of phylogenetics and bioinformatics as applied to the question of the root of the tree of life. An important aspect is the discussion of horizontal transfer, how this could affect the search for the root, and the issue of whether informational genes could potentially transfer between lineages as readily as operational genes are suggested to.

2. Woese CR, Kandler O, Wheelis ML: **Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eukarya.** *Proc Natl Acad Sci USA* 1990, **87**:4576-4579.

3. Jain R, Rivera MC, Lake JA: **Horizontal transfer among genomes:** •• **the complexity hypothesis.** *Proc Natl Acad Sci USA* 1999, **96**:3801-3806.

The authors argue for extensive gene transfer between prokaryotes during evolution and that it is genes of the operational class that are transferred most frequently. They suggest that transfer of informational genes is hindered by the many intermolecular interactions in which these macromolecules are involved. Informational genes include those for transcription, translation, replication, and GTPases. Operational genes are those for nearly all of metabolism, including regulation.

4. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, • Hickey EK, Peterson JD, Nelson WC, Ketchum KA *et al.*: **Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*.** *Nature* 1999, **399**:323-329.

Another whole microbial genome sequence from TIGR. This paper considers especially the issue of horizontal gene transfer, concluding on the basis of conserved gene order that some horizontal transfer occurs between eubacteria and archaea.

5. Aravind L, Tatusov RL, Wolf YI, Walker DR, Koonin EV: **Evidence for** • **massive gene exchange between archaeal and bacterial hyperthermophiles.** *Trends Genet* 1998, **14**:442-444.

Describes evidence that horizontal transfer from hyperthermophilic archaea to hyperthermophilic bacteria occurs more readily than to mesophilic bacteria. The authors conclude that this transfer may have been the defining event in the origin of hyperthermophilic bacteria.

6. Rivera MC, Jain R, Moore JE, Lake JA: **Genomic evidence for two** •• **functionally distinct gene classes.** *Proc Natl Acad Sci USA* 1998, **95**:6239-6244.

Using whole-genome data, the authors class genes as either operational or informational on the basis of function and demonstrate that the operational gene sets of bacteria and eukaryotes are more closely related than that of the archaea, whereas the archaea–eukaryote grouping holds for the informational gene set.

7. Snel B, Bork P, Huynen MA: **Genome phylogeny based on gene** •• **content.** *Nat Genet* 1999, **21**:108-110.

The authors use the gene content of 13 completely sequenced genomes for reconstructing the tree of life and rooting it. Unlike sequence-based phylogenies, the tree is built by examining similarities and differences in gene content, so that the presence or absence of a gene is counted as a character. The authors conclude that massive horizontal transfer events between distant groups is not supported by their results, and that their data largely support the 16S rRNA tree topology for the 13 genomes.

8. Woese CR: **The last universal common ancestor.** *Proc Natl Acad Sci USA* 1998, **95**:6854-6859.

9. Iwabe N, Kuma K-I, Hasegawa M, Osawa S, Miyata T: **Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes.** *Proc Natl Acad Sci USA* 1989, **86**:9355-9359.

10. Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF, Poole RJ, Date T, Oshima T *et al.*: **Evolution of the vacuolar H^+-ATPase: implications for the origin of eukaryotes.** *Proc Natl Acad Sci USA* 1989, **86**:6661-6665.

11. Brinkmann H, Philippe H: **Archaea sister-group of Bacteria?** •• **Indications from tree reconstruction artifacts in ancient phylogenies.** *Mol Biol Evol* 1999, **16**:817-825.

Uses a new method which applies the 'covarion' model to building a tree of life from signal recognition particle proteins. The authors conclude that the root is in the eukaryote branch and that earlier trees built with these sequences placing eubacteria as basal were as a result of long-branch attraction.

12. Lopez P, Forterre P, Philippe H: **The root of the tree of life in light of**
•• **the covarion model.** *J Mol Evol* 1999, **49**:496-508.
A description of a new method for applying the 'covarion' model to tree building that allows sites to alter their rate of evolution as secondary and tertiary structure evolves. Applied to the rooting of the tree of life, and with elongation factors, the authors conclude that the eubacteria are evolving at a higher rate than either archaea or eukaryotes, accounting for their basal position in earlier trees.

13. Penny D, Foulds LR, Hendy MD: **Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences.** *Nature* 1982 **297**:197-200.

14. Philippe H, Laurent J: **How good are deep phylogenetic trees?** *Curr*
•• *Opin Genet Dev* 1998, **8**:616-623.
A brief appraisal of the problems associated with building phylogenetic trees for deep divergences. Problems that are often ignored such as differences in evolutionary rate, sequence saturation, and fast-evolving lineages are addressed, and limitations of existing methods as well as possible solutions are described.

15. Teichmann SA, Mitchison G: **Is there a phylogenetic signal in**
•• **prokaryote proteins?** *J Mol Evol* 1999, **49**:98–107.
Using a data set of 32 proteins, it is shown that one gene which has undergone horizontal transfer can heavily influence the construction of a phylogenetic tree, even for a data set of 32 proteins. Upon removal of the offending gene, the remainder of the data set contained little information.

16. Forterre P, Philippe H: **Where is the root of the universal tree of**
•• **life?** *Bioessays* 1999, **21**:871-879.
An overview of problems associated with rooting the tree of life, along with arguments favouring prokaryotes being derived from a mesophilic ancestor that was eukaryote-like in many respects. The paper also reviews details of how eukaryotes and prokaryotes have arisen from such an ancestor.

17. Lockhart PJ, Steel MA, Barbrook AC, Huson DH, Howe CJ:
• **A covariotide model describes the evolution of oxygenic photosynthesis.** *Mol Biol Evol* 1998 **15**:1183-1188.
A mathematical test is introduced that can detect some cases where sites in a sequence are evolving under different constraints in different parts of the tree. It is well known in structural biology that two- and three-dimensional structrures of macromoles evolve over time (as predicted under W Fitch's covarion model); however, standard tree-building methods assume a site is always under the same contrstraints.

18. Benner SA, Ellington AD, Tauer A: **Modern metabolism as a palimpsest of the RNA world.** *Proc Natl Acad Sci USA* 1989, **86**:7054-7058.

19. Jeffares DC, Poole AM, Penny D: **Relics from the RNA world.** *J Mol*
•• *Evol* 1998, **46**:18-36.
An updated model of the RNA-world, describing arguments, both old and new, for the placing of various RNAs in the RNA-world, what gaps there are in our present understanding of this period, plus a review of ancient genome architecture from the viewpoint of information theory. The paper also describes a novel way of viewing the evolutionary transition from RNA to protein catalysts and explains why some RNAs have persisted while others have not.

20. Poole AM, Jeffares DC, Penny D**: The path from the RNA world.**
•• *J Mol Evol* 1998, **46**:1-17.
Here we attempted to establish what was known about the evolutionary transitions in going from an RNA-world to the emergence of the three domains of life. Included is a discussion of the origins of protein synthesis, the first proteins, messenger RNA, as well as aspects of the origins of DNA. Notably, we put forth a new hypothesis on the origin of introns, which we call 'introns-first'. Also discussed is the validity of using RNA-world relics for 'rooting' the tree of life. We conclude that the data we assemble are incompatible with a prokaryote-like Last Universal Common Ancestor.

21. Poole AM, Jeffares DC, Penny D: **Early evolution: prokaryotes, the**
•• **new kids on the block.** *Bioessays* 1999, **21**:880-889.
We argue for re-evaluating the nature of the Last Universal Common Ancestor. Emphasises the importance of continuity of function in evolution, and suggests that our understanding of the RNA-world and the Last Universal Common Ancestor should be mututally compatible. It proposes a feedback process (the Darwin-Eigen cycle) where improved accuracy of replication permits a larger genome size, which permits coding for more features, which permit more accurate replication.

22. Reanney DC: **Genetic error and genome design.** *Trends Genet* 1986 **2**:41-46.

23. Reanney DC: **Genetic error and genome design.** *Cold Spring Harb Symp Quant Biol* 1987, **52**:751-757.

24. Forterre P: **Thermoreduction, a hypothesis for the origin of prokaryotes.** *CR Acad Sci Paris III* 1995, **318**:415-422.

25. Hiller R, Zhou ZH, Adams MW, Englander SW: **Stability and dynamics in a hyperthermophilic protein with melting temperature close to 200 degrees C.** *Proc Natl Acad Sci USA* 1997, **94**:11329-11332.

26. Marguet E, Forterre P: **DNA stability at temperatures typical for thermophiles.** *Nucleic Acids Res* 1994, **22**:1681-1686

27. Greenstein JP, Winitz M: **Glutamic acid and glutamine.** *In Chemistry of the Amino Acids.* New York: John Wiley and Sons; 1961:1929-1954.

28. Legrain C, Demarez M, Glansdorff N, Piérard A: **Ammonia-dependent synthesis and metabolic channelling of carbamoyl phosphate in the hyperthermophilic archaeon *Pyrococcus furiosis*.** *Microbiology* 1995, **141**:1093-1099.

29. Galtier N, Tourasse N, Gouy M: **A nonhyperthermophilic common**
•• **ancestor to extant life forms.** Science 1999, **283**:220-221.
The authors compare the GC content of modern organisms in order to understand more on the nature of the LUCA and concludes that the ancestral GC content was too low for it to have been hyperthermophilic. A similar ancestral GC content was found using only the thermophilic organisms in the dataset.

30. Levy M, Miller SL: **The stability of the RNA bases: implications for**
• **the origin of life.** *Proc Natl Acad Sci USA* 1998, **95**:7933-7938.
The half-lives of RNA bases at temperatures characteristic of hyperthermohiles is shown to be too rapid for bases to accumulate in a prebiotic world. The authors conclude that life must originate at low temperatures or that theories for the high temperature origin of life must exclude the four bases in RNA.

31. Brocks JJ, Logan GA, Buick R, Summons RE: **Archean molecular**
•• **fossils and the early rise of eukaryotes.** *Science* 1999 **285**:1033-1036.
Identification of molecular biomarkers in 2700-million-year-old Archaean shales in Australia argues for the presence of photosynthetic organisms hundreds of millions of years before the atmosphere became oxidising. Perhaps more strikingly, the research also points to the presence of eukaryotes at this time, pushing back the earliest identification of these organisms by 600 million years.

32. Mushegian AR, Koonin EV: **A minimal gene set for cellular life derived by comparison of complete bacterial genomes.** *Proc Natl Acad Sci USA* 1996, **93**:10268-10273.

33. Becerra A, Islas S, Leguina JI, Silva E, Lazcano A: **Polyphyletic gene losses can bias backtrack characterizations of the cenancestor.** *J Mol Evol* 1997, **45**:115-117 [Mushegian AR, Koonin EV: **Response.** *J Mol Evol* 1997, **45**:117-118.]

34. Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Pontén T,
•• Alsmark UCM, Podowski RM, Näslund AK, Eriksson A-S,Winkler HH, Kurland CG: **The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.** *Nature* 1998, **396**:133-140.
The first complete genome for an α proteobacterium and thereby among the closest ancestors to mitochondria. The work demonstrates the effects of Müller's ratchet (accumulation of deleterious alleles in the absence of recombination) on the evolution of intracellular obligate microbes.

35. Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M,
•• Kowallik KV: **Gene transfer to the nucleus and the evolution of chloroplasts.** *Nature* 1998, **393**:162-165
Using whole genomes, mostly of chloroplasts, the paper demonstrates patterns of gene loss and of gene transfer to the nucleus. They find independent gene losses in multiple lineages and identify a large set (44) of chloroplast genes which had transferred from chloroplast to nucleus. It is concluded that gene loss and transfer in organelles is best explained in terms of Müller's ratchet.

36. Koonin EV, Mushegian AR, Bork P: **Non-orthologous gene replacement.** *Trends Genet* 1996, **12**:334-336.

37. Doolittle WF: **You are what you eat: a gene transfer ratchet could**
• **account for bacterial genes in eukaryotic nuclear genomes.** *Trends Genet* 1998 **14**:307-311.
The author proposes a mechanism for gene transfer from bacteria to the eukaryote nucleus. The model attempts to account for the apparent chimeric makeup of the nuclear genome and is a plausible mechanism by which eukaryotes could acquire genetic information from the diet.

38. Martin W: **Mosaic bacterial chromosomes: a challenge en route to**
•• **a tree of genomes.** *Bioessays* 1999, **21**:99-104.
A broad review of current knowledge on horizontal transfer. The author describes some interesting consequences of horizontal transfer with respect to phylogenetic analyses. Additionally, the likely differences between prokaryote–prokaryote transfer and prokaryote–eukaryote transfer are discussed.

39. Lynch M: **Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes.** *Mol Biol Evol* 1997, **14**:914-925.

40. Reanney DC: **On the origin of prokaryotes.** *J Theor Biol* 1974, **48**:243-251.

41. Pennisi E: **Genome data shake the tree of life.** *Science* 1998, **280**:672-673.

42. Pennisi E: **Is it time to uproot the tree of life?** *Science* 1999, **284**:1305-1307.