J Mol Evol (1998) 46:1–17

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 1998

The Path from the RNA World

Anthony M. Poole,* Daniel C. Jeffares, David Penny

Institute of Molecular Biosciences, Massey University, PO Box 11222, Palmerston North, New Zealand

Received: 14 January 1997 / Accepted: 19 May 1997

Abstract. We describe a sequential (step by step) Darwinian model for the evolution of life from the late stages of the RNA world through to the emergence of eukaryotes and prokaryotes. The starting point is our model, derived from current RNA activity, of the RNA world just prior to the advent of genetically-encoded protein synthesis. By focusing on the function of the protoribosome we develop a plausible model for the evolution of a protein-synthesizing ribosome from a high-fidelity RNA polymerase that incorporated triplets of oligonucleotides. With the standard assumption that during the evolution of enzymatic activity, catalysis is transferred from RNA → RNP → protein, the first proteins in the ''breakthrough organism'' (the first to have encoded protein synthesis) would be nonspecific chaperone-like proteins rather than catalytic. Moreover, because some RNA molecules that pre-date protein synthesis under this model now occur as introns in some of the very earliest proteins, the model predicts these particular introns are older than the exons surrounding them, the ''introns-first'' theory. Many features of the model for the genome organization in the final RNA world ribo-organism are more prevalent in the eukaryotic genome and we suggest that the prokaryotic genome organization (a single, circular genome with one center of replication) was derived from a ''eukaryotic-like'' genome organization (a fragmented linear genome with multiple centers of replication). The steps from the proposed ribo-organism RNA genome → eukaryotic-like DNA genome → prokaryotic-like DNA genome are all relatively straightforward, whereas the transition prokaryotic-like genome → eukaryotic-like genome appears impossible under a Darwinian mechanism of evolution, given the assumption of the transition RNA → RNP → protein. A likely molecular mechanism, ''plasmid transfer,'' is available for the origin of prokaryotic-type genomes from an eukaryotic-like architecture. Under this model prokaryotes are considered specialized and derived with reduced dependence on ssRNA biochemistry. A functional explanation is that prokaryote ancestors underwent selection for thermophily (high temperature) and/or for rapid reproduction (*r* selection) at least once in their history.

Key words: Genome structure — Introns — Molecular evolution — Molecular fossils — Origin of translation — Prokaryote origins — *r* Selection — Theoretical biology — Thermoreduction — Tree of life

Introduction

Under a Darwinian model the evolution of life involves a continuous series of ancestors with a large number of intermediate stages, all of which need to be functional. Of these stages, the evolution of encoded protein biosynthesis is one of the major problems in developing a precise theory for the origin of life. The evolution of protein biosynthesis demarcates the beginning of modern biochemistry, and hence also modern life, and we will refer to this stage as the ''breakthrough organism.'' The assumption is made that the breakthrough organism arose from a population of ribo-organisms that utilized RNA as both genetic material and catalyst. Again under a Dar-

**Present address:* Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan
Correspondence to: D. Penny; e-mail D.Penny@massey.ac.nz

winian mechanism, a complex structure such as the ribosome could not just arise de novo, so it is essential to identify the function of the protoribosome and address how it could have been co-opted or recruited into encoded protein biosynthesis.

This stage would have been followed by the evolution of many new structural and catalytic proteins before a more complex organism developed that was the Last Universal Common Ancestor (LUCA) of all modern life. Our approach is to examine those RNAs that have survived from the last ribo-organism until the present day (Jeffares et al. 1997). This analysis now leads to the interesting conclusion that an encoded translation system could develop by numerous small steps and that the genome organization of the last universal common ancestor had many features considered characteristic of eukaryotic organisms. On the basis of inferred molecular fossils considered here and elsewhere (Jeffares et al. 1997), we develop a model describing the path from the RNA world, which includes discussions on the origins of introns, mRNA, the first proteins, and the likely structure of the genes in which they were housed.

In discussions of the origin of prokaryotes and eukaryotes, it is commonly assumed that prokaryotes predate eukaryotes on the basis of their apparent simplicity, and a number of phylogenetic studies appear to support this conclusion (see Doolittle 1995 for a summary). However, the reliability of such tree-building methods for resolving such deep divergences is subject to debate (Doolittle 1995; Baldauf et al. 1996; Lockhart et al. 1996), and it has been demonstrated that models used in current tree-building methods cannot yet give accurate results even for early photosynthetic relationships (Lockhart et al. 1996, see later). Given the predictive power of the RNA world theory (Forterre 1995b, 1996; Jeffares et al. 1995, 1997) and the relevance of the model described here for the path from the RNA world, an alternative method by which to address the nature of the last universal common ancestor is to consider the wealth of metabolic data, or molecular fossils, currently available. We consider the RNA relics in contemporary metabolism as remnants from the breakthrough organism (Jeffares et al. 1997), pre-dating the last universal common ancestor. Thus such relics comprise an alternative outgroup for rooting the tree of life.

In this article, the word ‘‘genome’’ and the phrase ‘‘genome organization’’ are used only to indicate whether the genome is circular or linear, or fragmented or continuous; whether there are single or multiple centers of replication; or to refer to the nature of the genetic material and the genome copy number. Consequently, the use of the words prokaryotic and eukaryotic in relation to genome organization refer only to these features (Fig. 1). The possible presence or absence of a nucleus (or other cellular compartmentation) and the possible use of histones in chromosome packaging at this stage in

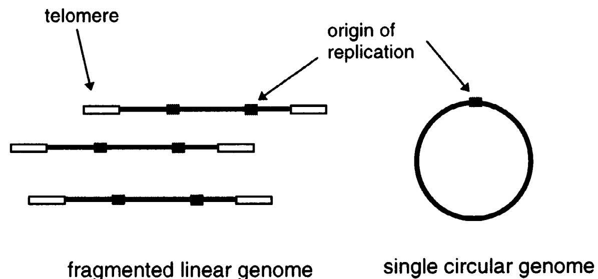


Fig. 1. The two main extant genome organizations. Genome organization only includes information such as size of the genome; linear or circular; continuous or fragmented; copy number; presence or absence of intervening sequences; single or multiple centers of replication. It does not include cytological information such as cellular or acellular; membranes present or absent, or details of cellular compartmentation (such as a nuclear structure). As such, it is possible for an organism to lack a nucleus but still have a ‘‘eukaryotic genome organization.’’ It is not clear yet whether a single origin of replication occurs in archaea as well as eubacteria (Bult et al. 1996). For this reason we do not preclude the existence of multiple origins of replication within the archaea; however, until more information comes to light the model is based on the better-studied system of eubacteria.

the evolution of life are not considered. Not all prokaryotes have a circular genome (Hinnebusch and Tilly 1993), but in this article, all references to ‘‘prokaryotic genome organization’’ should be taken as meaning a covalently closed circular genome composed of double-stranded DNA. Because archaea and eubacteria are fundamentally similar in genome organization (Baumann et al. 1995) we distinguish between them only as necessary. This is not to say that prokaryotes split from eukaryotes as a single group which only later split to form eubacteria and archaea; under the thermoreduction hypothesis (Forterre 1995a, 1996) and the plasmid-transfer model (see later), a prokaryotic-type genome can conceivably have arisen more than once.

An examination of the genome organization of the three broad domains of life (archaea, eubacteria, and eukaryotes) leads to a testable model describing the molecular mechanism by which a prokaryotic-like genome architecture could have arisen from the proposed genomic structure of the LUCA. This plasmid-transfer model proposes that, by a process of reverse transcription, the genetic information housed on the linear, fragmented genome of the LUCA was transferred to a circular plasmid-like molecule, thereby producing the prototype prokaryote genome organization. The metabolism of the LUCA, like that of modern eukaryotes, is expected to have been heavily dependent on RNA, and the model also offers an explanation as to how many of the RNA processing events of eukaryote metabolism could be eliminated from an emerging prokaryotic lineage concurrent with genome circularization. Our conclusion is that prokaryote genome structure is derived, all prokaryotes having undergone *r* selection and/or a thermophilic stage to produce a smaller, compact, and efficiently organized genome.

Table 1. Features required in a protoribosome[a]

	Function in the RNA world	Without tags	With amino acid tags
1	A large complex structure must have a function to evolve or be maintained by selection	Yes	Yes
2	Existence of, and role for, tRNA-like molecules	Yes	Yes
3	An anticodon on the tRNA (for adding to growing ssRNA)	Yes	Yes
4	A mechanism for charging a tRNA with a specific amino acid	—	Yes
5	A ribosome precursor consisting of two polynucleotides (functionally equivalent to contemporary rRNA species)	Yes	Yes
6	ssRNA (equivalent to messenger RNA in the modern world)	Yes	Yes
7	A recognition site on the ribosome for ssRNA	Yes	Yes
8	A recognition site on the ribosome for tRNA that allows the anticodon to react with the ssRNA (decoding)	Yes	Yes
9	A fast synthetic reaction that is completed within the time the anticodon and ssRNA bind (before they separate by diffusion)	Yes[b]	Yes[b]
10	A ratchet mechanism to move the ssRNA through the ribosome by the length of the anticodon	Yes	Yes
11	A one-to-one relationship between the anticodon and amino acids (the triplet code)	—	Yes

[a] The likely presence of the features is indicated under the simple model (without amino acid tags on the tRNAs) and in the full model (with tags)
[b] The reaction carried out would necessarily be different in the protoribosome

A Path from the RNA World

Although the first test is its plausibility, a model is much more useful if further hypotheses and/or tests can be developed from it. The model of the last ribo-organism described in Jeffares et al. (1997) leads to inferences concerning later stages of evolution—the origin of protein synthesis, the development of a DNA–protein world, and then differentiation into prokaryote and eukaryote genome organizations, and these problems are discussed in turn.

A Model for the Origin of the Ribosome and Protein Synthesis

The apparent problem with developing a templated protein synthetic machinery is that many partial processes are necessary and all must be established before genetically encoded protein synthesis can function. Because of the importance of this point we enumerate 11 processes in Table 1. The first point is that all large complex structures, such as a ribosome precursor in a ribo-organism, must have an essential function both to evolve and to be maintained by the processes of natural selection. In the absence of selection, and with a high error rate of RNA replication, the protoribosome would decay over comparatively few generations. Thus one of the most critical steps in the origin of protein synthesis is to explain the function of the protoribosome prior to its recruitment into protein synthesis. There are many places in molecular evolution where, for example, an enzyme gets recruited into a new function, but de novo origin is uncommon and is not an option for a complex structure such as a ribosome.

It is not reasonable for a model to assume that ‘‘all these functions (Table 1) just happen to coincide’’—there must be an explicit mechanism that allows each step to develop sequentially. In the early stages of an RNA world it is assumed that, because of the limited replication accuracy, RNA molecules would not exceed a few hundred bases (see Eigen 1992). Larger molecules could then arise later as replication increased in accuracy. It is possible that the several active sites of modern ribosomes evolved as separate ribozymes, to be joined by recombination once replication fidelity could reliably produce entire rRNAs. Small RNAs could thus have acted in *trans* to form a functioning ribosome. Possible relics of this history are that decoding (the interaction of tRNA anticodons with the ribosome) can be mimicked by a small RNA analog of the rRNA region thought to be involved with decoding in intact ribosomes (Purohit and Stern 1994), and the finding that the α-sarcin loop appears to be a modular RNA (Szewczak and Moore 1995). The general problem is similar to the origin of sexual reproduction and meiosis (Penny 1985). Darlington (1958) had claimed that no Darwinian mechanism was possible for the evolution of a process as complex as meiosis because so many steps were apparently necessary before it would confer benefit to the organism. A model was demonstrated (Penny 1985) where each step could evolve sequentially. Similarly, we show here that intermediates are possible for all the steps in the origin of protein synthesis.

One possible model for the origin of template-directed protein synthesis is a ribosome precursor that was an RNA polymerase—specifically, one that adds trinucleotides to the growing RNA molecule (Fig. 2; see Weiss and Cherry 1993; Gordon 1995). Consider a tRNA-like

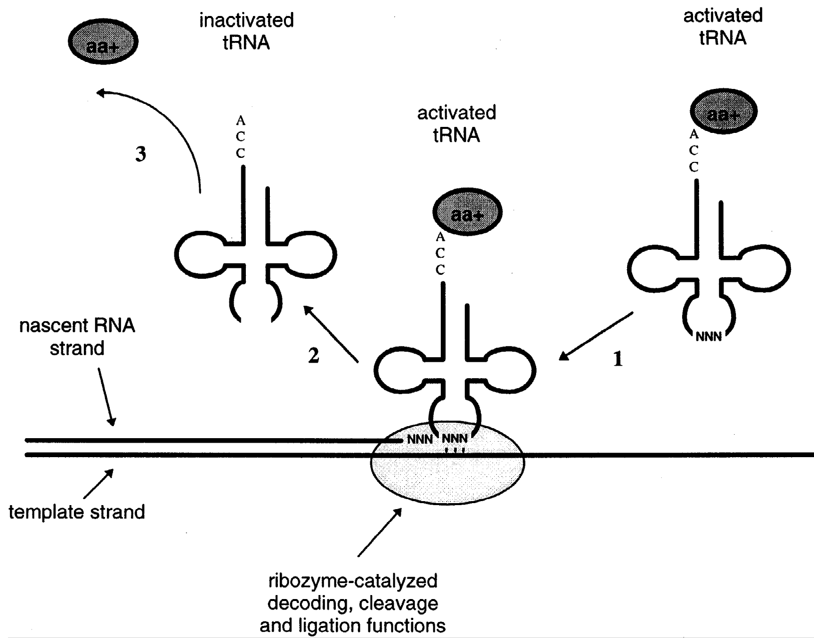


Fig. 2. An ancient RNA replicase as the precursor of the ribosome. The modern ribosome contains RNA and a large number of proteins, but its origins were undoubtedly in the RNA world. The figure shows a possible model for the origin of the ribosome from an RNA replicase/polymerase that adds triplets to a growing RNA. (*1*) A positively charged amino acid tag helps the replicase recognize the tRNA, bringing them into contact. (*2*) The anticodon triplet is added to the growing chain by a process of cleavage and ligation similar to that catalyzed by the modern spliceosome. (*3*) The 23S rRNA cleaves the positively charged amino acid from the acceptor stem, and the used tRNA is released. The stage is then set for the origin of peptide bond formation, driven thermodynamically by aminoacyl tRNA cleavage.

molecule that is charged with a trinucleotide at the position of the present anticodon; if the trinucleotide is complementary to the next three nucleotides on the ssRNA being copied it could be incorporated into the new RNA. Several authors have suggested short oligonucleotides could be used for RNA synthesis in an RNA world (Sharp 1985; Orgel 1986; Doudna and Szostak 1989; Gordon 1995). An advantage of adding short nucleotide chains, rather than single nucleotides, is that they would H-bond longer to the RNA template, giving the polymerase more time to join the short chain by a transesterification reaction.

Allowing a longer time for reactions to occur is expected to be important for an RNA-catalyzed mechanism that, compared to protein catalysts, reacts more slowly (turnover times are in the order of minutes, Table 1 in Jeffares et al. 1997). Although modern polymerase enzymes require only a single nucleotide pairing to guarantee specificity (Switzer et al. 1989; Piccirilli et al. 1990) it is expected that lower turnover times for ribozymes would be too slow for a high-fidelity RNA polymerase. Eigen and Schuster (1978) report that for five AT pairs the association time, before diffusion separates them, would be only milliseconds, or up to a few seconds with five GC pairs. We consider that the slow rate of reaction was a limiting feature for the accuracy of RNA synthesis by ribozymes. Experimental support for this analysis comes from Ekland and Bartel (1996), who report that a ribozyme derived by artificial (in vitro) evolution can indeed catalyze the addition of single nucleotides from triphosphates. However, the accuracy is relatively low, more than one error per 100 nucleotides, and the rate of addition is about seven reactions per hour.

A possible reason for a triplet, as opposed to shorter or longer oligonucleotides, arises from this same paper (Ekland and Bartel 1996), which reports that ribozyme reactions are very slow after adding two or three nucleotides. This may be related to the distance a ribozyme can extend and still carry out the reaction. After the addition of three nucleotides a dissociation/reassociation reaction of ribozyme and substrate may be necessary, or a mechanism for moving the ribozyme three nucleotides along the RNA template may be needed. This second alternative could be the origin of the ratchet mechanism that moves the ribosome three nucleotides along the mRNA—requirement number 10 in Table 1. Thus the length of the codon (a triplet) may already have been established in the RNA world. A similar periodicity, in this case after adding six nucleotides by the RNP telomerase, occurs in *Tetrahymena* telomere synthesis in vitro (Collins and Greider 1993). Overall, the results of such an RNA polymerase (Ekland and Bartel 1996) support our general analysis of the need for, and the problems of, a high-accuracy RNA polymerase in the final stages of the RNA world.

There may be other reasons militating against longer oligonucleotides, in spite of the longer time available for a reaction to take place. The number of possible oligonucleotide substrates increases exponentially with length—but it is expected to take four times longer to find the right match for a tetranucleotide than for a trinucleotide. Stability increases linearly, the number of possibilities exponentially. There is a tradeoff between increasing accuracy and slower replication rates as longer oligonucleotides are considered. In addition, accuracy could be increased by additional recognition sites (tags).

Increased replication accuracy could occur if an amino acid tag occurred on the pre-tRNA with a code already established in the RNA world (Nagel and

Doolittle 1995; Wetzel 1995; Härtlein and Cusack 1995), that is, before proteins had the main catalytic role. The relationship between amino acid and anticodon could have been established with an amino acid attached to the CCA of the pre-tRNA, thereby increasing accuracy of the RNA polymerase. This approach is favored by Taylor and Coates (1989) and Maynard Smith and Szathmáry (1995, p81ff), particularly as there are regularities between position of the codon and the size, biosynthesis, and polarity of the amino acids they encode. Such an amino acid tag would not initially have been involved in protein synthesis but could have increased the specificity of a preribosomal RNA polymerase, an improvement over just using the trinucleotide for specificity. A difficulty is that we would not expect a different amino acid for each of 64 triplets, so under this model some redundancy in the amino acid triplet code would already exist in the RNA world. It is even possible that the RNY of Crick (1968) and Eigen and Winkler-Oswatitsch (1981) could have existed, increasing the accuracy of RNA replication by helping maintain the triplet reading frame. A further possibility is that the amino acids were more than ''tags'' and were involved, for example, by being hydrolyzed from the tRNA and driving the reaction that incorporated the triplet. These two extensions to the basic model for the origin of a protein-synthesizing ribosome are more speculative, though they would solve step 11 (Table 1) of the series of necessary stages in the evolution of protein synthesis and/or involve the amino acids in metabolism from a very early stage.

Maizels and Weiner (1987, 1994) point out that early tRNA molecules may have consisted of only part of the current tRNA molecule. The likelihood of this is supported by the demonstration that partial tRNA molecules can be charged with their appropriate amino acid (Schimmel and de Pouplana 1995). Several authors (Keese and Gibbs 1992; Maizels and Weiner 1994) have suggested that initially a positively charged amino acid, or short peptide, would neutralize negative charges on RNA, allowing a more tightly packed tertiary structure. With regard to RNA-mediated charging of tRNA, Illangasekare et al. (1995) have succeeded in evolving in vitro an RNA capable of performing this task.

Several versions of the model are possible regarding the interaction of the charged tRNA and the replicase in terms of the ratchet mechanism (requirement 10 in Table 1). Assuming that the positive charge on the amino acid is involved in binding of the aminoacylated tRNA to the replicase complex (Fig. 2, step 1), cleavage of the amino acid from the tRNA (Fig. 2, step 3) would then allow release of the tRNA. Affinity of the replicase for activated tRNA could cause a conformational change that releases the used tRNA and allows binding of the incoming activated tRNA; this might be envisaged as being carried out by the 23s rRNA. Hence, the model ''with tags'' (Table 1, Fig. 2) allows possible refinements to the ratchet mechanism as well as to tRNA binding and complex stability.

If such a protoribosome (an RNA polymerase involved in either replication or transcription) was the antecedent of modern 18S, 28S, and 5.8S rRNAs then all of the steps listed in Table 1 except number 4 are feasible (though some features such as the triplet would have a different function). The process would bind both ssRNA and tRNA precursor in the correct position for the anticodon, move the ssRNA three nucleotides after every cycle, and recognize control sequences for initiating and terminating polymerase activity. Because of the Eigen limit on genome size (Eigen 1993), we expect there to be very strong selection for increased fidelity of RNA synthesis in the RNA world.

The Origin of mRNAs and Introns First

A ssRNA molecule that became mRNA must have been present in ribo-organisms in some other role before the evolution of translation. The origin of the information in the mRNAs is perhaps the most difficult problem to resolve because we would not expect these ribo-organisms to contain meaningful information about future protein sequences. So far we have not distinguished between the protoribosome being involved in replicating RNA genomes, and transcribing active RNA enzymes from the genome. However there are more similarities between transcription of a single ribogene and translation of a single gene; it is in the transcription of ''ribogenes'' that we consider translation first arose.

In our model of the last ribo-organism (*Riborgis eigensis*) there are many RNA-processing steps, including cleavage and splicing of transcripts that end up as ribozymes (Jeffares et al. 1997). mRNAs may have arisen as byproducts of these ribozyme processing reactions, and it is from the unused genetic material between these ribozymes that mRNAs arose (Fig. 3). We suggest that intronic small nucleolar RNAs (snoRNAs) show examples of the spacers between ribozymes that gave rise to mRNA. These spacers are now exons (Fig. 3).

Small nucleolar RNAs (snoRNAs) are often found encoded within intronic regions of ribosomal and heat-shock proteins (Fig. 3; reviewed in Maxwell and Fournier 1995). We have concluded, based on the evolutionary trend RNA → RNP → protein, that these snoRNA molecules pre-date the origin of protein translation (Jeffares et al. 1995) and therefore predate the exons surrounding them. This we call the ''introns-first'' theory; contemporary introns housing functional RNAs are relics of the RNA world genome organization, and the newer protein regions surrounding them represent sequences that were originally noncoding and from which protein genes were eventually spawned (Fig. 3). This is consistent with our model for the origin of protein translation, because the existing RNA genes are not disrupted by the advent of new protein-coding genes.

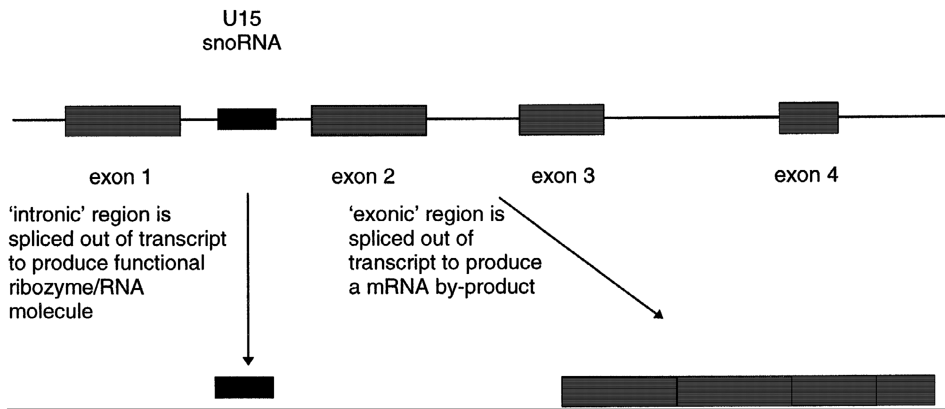


Fig. 3. Ancient ribozymes surviving in the introns of ancient proteins? The structure of human ribosomal protein S3 (Tycowski et al. 1993). The snoRNA U15 is encoded in an intron. Intron-encoded snoRNAs are common for those snoRNAs that contain sequences complementary to rRNA (antisense snoRNAs). We propose that intronic snoRNAs, such as U15 (*solid block*), date back to the RNA world, pre-dating the exons (*shaded blocks*) that surround them, these ''exonic'' regions initially being noncoding and metabolically nonfunctional. The first mRNAs arose as byproducts of ribozyme excision reactions.

A reaction that produces molecules like this in extant organisms can be found in the production of snoRNA (small nucleolar RNA) from introns of many chaperone-like proteins (Jeffares et al. 1997). An unusual transcript is found both in mouse and human (Tycowski et al. 1996) where the ''exons'' do not encode a protein but are spliced together to produce an mRNA that is degraded without being translated. However the intronic regions of the gene encode eight snoRNAs that are conserved between mouse and human, displaying a conserved ''U-turn'' motif exhibiting common complementarity to rRNA. This is a case where the intronic regions are important and the exonic regions are not. It could be argued that this architecture might be the result of recent loss of protein function with the retention of the functionally important snoRNAs, and not necessarily an example of a bona fide relic of the architecture of *R. eigensis.* Nevertheless, this fragmented architecture highlights two important points: such an architecture is possible, and function may evolve in fragments.

The genome is dynamic with respect to the position of genes over time, so the snoRNA genes may well have moved. However, if this occurred extensively we might expect insertion into some more recent genes, such as catalytic genes. Whether this proves to be the case or not, the occurrence of snoRNAs almost exclusively in the introns of chaperone-like proteins is more easily explained by suggesting they predate these proteins and that the common genomic location of the two types of molecule is the result of these early proteins arising in the noncoding regions separating snoRNA genes. An alternative is that this gene arrangement is not ancestral but evolved secondarily, selected, for example, because it allows stoichiometric co-expression of these RNAs and their host proteins. However, non–intron-encoded snoRNAs have been found, and because one mRNA can yield many proteins but there is only a single snoRNA molecule per intron, the stoichiometry is not one to one. At present there does not appear to be a selective advantage to the gene arrangement and we prefer the historical explanation. Given also that the snoRNAs had to be spliced out off the RNA template before proteins arose, this provides a means by which to justify a fragmented gene origin for at least some protein genes. The exon shuffling theory (Gilbert and Glynias 1993) is not the only possible explanation for the early origin of introns in protein genes.

Although we also consider the spliceosome to be old and to predate protein synthesis, we are neutral with respect to whether introns in general are ''early'' (Gilbert and Glynias 1993) or ''late'' (Palmer and Logsdon 1991). However, we do expect that the spliceosome predates protein synthesis; that introns in some heat-shock and ribosomal proteins pre-date the exons that surround them; and that consequently these exons are more recent than their introns. We cannot apply this conclusion to all introns, only to this limited number. The position of the universal root of the tree of life (see later) also affects the interpretation, and additional work is required to settle the more general question of introns early vs late. The fortuitous production of the first mRNAs may well have been an artifact of the spliceosome having coupled cleavage and ligation functions; it was only later that this functionless RNA was taken on as a source of protein genes. Furthermore, we predict that other such equivocal examples will be uncovered, requiring that both the concept of interrupted genes and their antiquity be reconsidered. It appears that this organization is likely to be more ancient than previously supposed.

A corollary of the introns-first theory is that it identifies some of the first proteins. The proteins that house intronic snoRNA genes seem to fit the category of phylogenetically ubiquitous chaperone-like proteins. The presence of snoRNA genes in the introns of ribosomal

proteins (reviewed in Séraphin 1993; Maxwell and Fournier 1995), HSC70 heat-shock cognate protein (Liu and Maxwell 1990; Leverette et al. 1992) and RCC1 (Kiss and Filipowicz 1993)—all chaperone-like under our definition—may be a further indication of their antiquity.

The Evolution of the First Protein-Coding Genes

Several aspects of the evolution of protein synthesis are more plausible if the first proteins were low-specificity RNA-binding proteins with chaperone-like activity, rather than requiring the first proteins to be catalysts as suggested, for example, by Jurka and Smith (1987). Chaperone-like proteins (as opposed to structural or catalytic proteins) are widespread. The classical concept of chaperone proteins (Lund 1995; Hartl 1996) requires that they do not stay bound to their substrate and this definition includes proteins involved in heat-shock response and protein folding and transport. However, in our concept of chaperone-like proteins we include proteins bound with RNA which are not themselves catalytic. There is a wide range of RNA-binding proteins (Wootton 1994; Draper 1995) and these include ribosomal proteins (though some such as L2 may now also be catalytic, Cooperman et al. 1995), spliceosomal Sm proteins, and the protein moieties of snoRNPs and RNase P. Extending the concept to proteins that stabilize RNA structure results in a wide range of chaperone-like proteins.

Two additional factors should be considered: the ease with which new novel sequences may have a useful function and the utility of ''low-complexity'' regions of proteins (those composed of a small number of amino acids). In a few cases there is evidence that some novel protein sequences may have a useful function. RNA viruses are under intense selection to use their genome as efficiently as possible because the lower fidelity of RNA replication places strong limits on the size of the genome (Reanney 1982). It is therefore not surprising that RNA viruses show a wide range of ''overlapping'' proteins where two or more proteins are read from the same nucleic acid sequence (Keese and Gibbs 1992) though in different reading frames. In several cases Keese and Gibbs (1992) have identified the original and the new protein, and they report that the original protein is probably catalytic and the new protein has a noncatalytic role. An important consequence is that new noncatalytic proteins are formed more easily than catalytic ones, and thus it is reasonable to postulate that the first proteins were noncatalytic and chaperone-like.

Wootton (1994) summarizes many recent reports of proteins having ''low complexity,'' that is, consisting of a small number of amino acids. Up to half of all new sequences reported have at least one ''low-complexity'' region, but in about a quarter of these, low-complexity regions form a significant portion of the molecule. These proteins are involved in several aspects of mRNA processing, transcription and its regulation, and binding to RNA and DNA. Until many more sequences are determined, it is not possible to decide whether particular low-complexity proteins are ancient. However, what they do show is that only a small number of amino acids are required to make proteins that are functionally important—and that many of these simple proteins interact with RNA. Such a conclusion supports the idea that the earliest genetically encoded polypeptides that interacted with RNA did so in an auxiliary role.

We suggest then that short, possibly positively charged, chaperone-like peptides in the RNA world would increase stability and help maintain ribozyme tertiary structure (thereby relaxing the need for high concentrations of metal ions such as Mg^{2+}). Such a function need not have been genetically encoded at the outset; the precise order of amino acids in such an early system may not be crucial; and, consequently, an early peptide synthetic system need not have required a high accuracy for synthesis. Peptides at this stage of evolution are unlikely to be included in a catalytic site. Once protein synthesis became genetically encoded, a means by which to consistently produce such peptides would be available, though the efficiency of the first ribosome is not expected to be high, limiting it to the production of low-complexity chaperones only. However, the increased stability conferred upon ribozymes (particularly those involved in genome replication and translation) by these new chaperone-like proteins would allow increasingly accurate translation to evolve and later, for proteins to develop sophisticated, catalytic functions.

This model then predicts that catalytic proteins only evolved after the accuracy of translation improved to the extent that longer and accurately synthesized proteins were possible. We envisage a positive feedback system where these early, gene-encoded short peptides increased the accuracy of peptide synthesis, via ribozyme stabilization, thereby permitting progressively more complex proteins requiring higher translation fidelity. With small substrates the ''chaperone-like'' protein may eventually take over the catalytic role (Maynard Smith and Szathmáry 1995, p81), though once some effective enzymes existed in a cell they could be duplicated and one copy could take over and replace another ribozyme.

Interestingly, a recent report (Wool 1996) emphasizes the multifunctional nature of ribosomal proteins. It appears that most of these proteins have functions additional to their role in the ribosome, and while it is likely that some may have been recruited to the ribosome from elsewhere in metabolism, given that we infer that these proteins are very ancient, we favor the opposite possibility—that many of these proteins were initially nonspecific chaperones of a sort, which acquired many central functions in the first period of the ''RNP world'' (the stage immediately after the evolution of translation).

Table 2. Molecular and cytological features of eukaryotes that need to be explained by any theory for the origin of eukaryotes[a]

		Endosymbiotic origin	Present theory
Molecular mechanisms			
	RNA in ribosome assembly and processing	No	Yes
	mRNA processing	No	Yes
	spliceosome	No[b]	Yes
	40+ small RNAs	No	Yes
	Slow processing times in eukaryotes	No	Yes
Cytological features			
	Nuclear membrane	Yes	No
	Nucleolus	No	No

[a] The endosymbiotic theory explains the origin of the nuclear membrane, but not the features of RNA metabolism in eukaryotes
[b] A post hoc explanation is the derivation of splicing from some form of transposable element. It is not, however, a prediction of the theory

Using the RNA World to Root the Tree of Life

Extant RNA molecules that have catalytic function can be considered fossil relics of the RNA world (Jeffares et al. 1997), a period in the evolution of life that predates the Last Universal Common Ancestor. Consequently, they can be used as an outgroup to root the tree of life. This is an alternative to using sequence data that have major problems for such deep divergences (Lockhart et al. 1996). We examine the compatibility of the data gleaned from the RNA-world model with rootings of the tree that place prokaryotes at its root. This method has the advantage that many of the hypotheses we propose are experimentally testable.

Central to the overall argument is that we see no short-term selective advantages, of the type required for Darwinian evolution, that could drive evolution from a prokaryotic to a eukaryotic genome organization. The following list, summarized in Table 2, demonstrates some of the observations that would not be expected with a prokaryotic origin for life. These observations are explained better under the alternative viewpoint that a eukaryote-like genome organization is ancestral (we return to this later under *r* Selection in Prokaryotes).

1. *Eukaryotic metabolism has many relics of the RNA world, while prokaryotes have fewer.*
   Eukaryotes contain spliceosomes, snorposomes, telomerase, vault RNAs, and self-splicing introns, all of which are absent from prokaryotes (see Fig. 2 of Jeffares et al. 1997). Given that proteins are catalytically superior to RNA, how can the heavily RNA-dependent RNA metabolism in eukaryotes be considered to postdate, let alone be derived from, the largely protein-utilizing metabolism of prokaryotes? As discussed in Jeffares et al. (1997), there seems to be no reasonable selection pressure that would favor the replacement of protein enzymes with turnover numbers of 10^3 to 10^6 with ribozymes with turnover numbers of about 1 (Table 1 of Jeffares et al. 1997). A prokaryotic model for the last universal common ancestor hence seems incompatible with the RNA world theory.
2. *mRNA and rRNA processing are fast and efficient in prokaryotes.*
   Ribosomal RNA processing in eubacteria begins while the polycistronic rRNA is still being copied (Morrissey and Tollervey 1995) but seems not to begin until the entire molecule is transcribed in eukaryotes (Steitz and Tycowski 1995). Similarly, mRNA processing and the time before a message is translated take much longer in eukaryotes. Many new RNAs, RNP particles, and proteins would need to evolve de novo to derive eukaryotic metabolism from prokaryotic. With rRNA processing in eukaryotes there are over 30 snoRNAs (small nucleolar RNAs) and their associated proteins involved (Maxwell and Fournier 1995; Tollervey 1996) that would not only have to appear de novo, but in addition *usurp* existing protein enzymes which were carrying out the process more efficiently (Jeffares et al. 1995).
3. *There is no selective advantage for the recent origin of mRNA splicing and the spliceosome in eukaryotes.*
   A Darwinian evolutionary model would require the *simultaneous* insertion of introns with the development of a complex RNA-containing splicing apparatus to remove them. What selective forces would favor a whole range of new RNA and proteins (the spliceosome) just to achieve mRNA production in 1 h instead of 1 min? In particular, it would be expected that protein enzymes, rather than ribozymes, would have been recruited from elsewhere in metabolism to carry out cleavage and ligation of RNA and be more efficient than a complex ribozyme ever could. The origin of the spliceosome is most simply explained as part of the metabolic complexity of the RNA world and a rationale for this is given in Jeffares et al. (1997). Note that such a model does not preclude the later evolution of new forms of mobile elements that could spread. The loss of splicing from prokaryotic lineages will be considered presently.

4. *Eukaryotic telomerases appear to be ancient.* What are the forces that led to ribonucleoprotein telomerase handling replication of linear chromosomes when replication was already handled adequately without the involvement of RNA in the circular genomes of prokaryotes? Though some prokaryotes have linear genomes with simple telomere structures akin to those found in linear DNA viruses (Hinnebusch and Tilly 1993), indicating that linearization of bacterial genomes is possible, it seems unlikely that the complex telomeres of eukaryotes are a recent addition to the eukaryotic genome architecture. The RNA component of the telomerase enzyme has at least a partial role in the overall catalytic function of the molecule (Gilley et al. 1995; Gilley and Blackburn 1996), and a number of features of both the protein and RNA components favor an early origin for both telomeres and telomerase (Maizels and Weiner 1993, 1994; Collins et al. 1995; Jeffares et al. 1997). Moreover, a recent derivation of circular chromosomes from linear can be explained both from the point of view of selective pressure and from a mechanistic viewpoint, and this will be discussed presently.
5. *Haploidy may be a derived trait of genomes.* Haploidy (in prokaryotes) and a single origin of replication (as in eubacteria) necessarily requires a high-fidelity replication apparatus, whereas a diploid or polyploid genome broken into a number of linear chromosomes, each with multiple origins of replication (as in eukaryotes), does not have such stringent requirements for high fidelity (Reanney 1974, 1987). Having only one copy of the genome removes the safety net that having two or more copies of each gene provides (Koch 1984; Reanney 1987); deleterious mutations in an essential gene would cripple the metabolism of an organism. Hence, haploidy is most likely an advanced feature that only became possible after an accurate replication apparatus became available. While haploidy could conceivably have been a feature of the last universal common ancestor (since it potentially had quite an advanced metabolism) we disregard this hypothesis on the grounds that the derivation of a circular genome from a linear fragmented genome has a foreseeable selective advantage under certain environmental conditions (see next section) and that many features of eukaryotic genome organization can be extrapolated back to the predicted genome organization of the last ribo-organism.

It has been suggested that the eukaryotic genome structure is derived from a linear dsRNA genome with simple telomeres and that its constituent chromosomes increased in length by the fusing of shorter replicative units (Szybalski and Szybalski 1974; Reanney 1974, 1979; Darnell 1981; Carlile 1980; Forterre 1992). Interestingly, the presence of silent origins of replication in yeast, and higher eukaryotes, might be due to streamlining of the replication system, as so many origins are not needed for replication. However, what proportion of these are functionally inactive or only temporally inactive is not clear (Fangman and Brewer 1991, for review).

Our model for the genome of the last ribo-organism (Jeffares et al. 1997) is consistent with the hypothesis we propose here—that the last universal common ancestor had a genome that was more eukaryote-like than prokaryote-like. In this scheme of genomic evolution, eukaryotic genomes do not change in basic architecture and organization significantly, although they become much larger and recruit proteins for essential functions—such as the two protein components of telomerase (Collins et al. 1995), proteins involved in the synaptonemal complex (Loidl 1991), and the histones that package chromatin. Because sophisticated mechanisms are required for reliable separation of sister chromosomes during cell division, and an increase in copy number does not hinder selection (Koch 1984), the tendency was for genetic material to *increase* in the early stages of cellular life, leading to a large, highly redundant (eukaryote-like) genome.

Evolution of a Prokaryotic Genome Organization

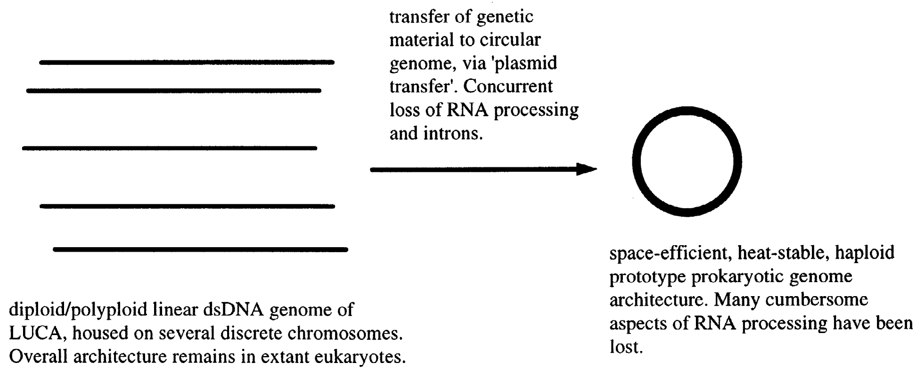
Deriving a prokaryotic genome organization from a eukaryotic organization is relatively straightforward in that both strong selective advantages (the thermoreduction hypothesis and/or *r* selection) and mechanisms for the simplification of processing (plasmid transfer model, next section) are available. Our conclusion (Jeffares et al. 1995) is that a fragmented, intron-containing, diploid and linear genome is ancestral, and that the streamlined single circular operon-containing genome (Carlile 1980, 1982) is derived from it (Fig. 4A). This derivation of a prokaryote genome is consistent with the transfer of catalysis from RNA → RNP → protein during evolution (Jeffares et al. 1997).

Retroviruses as a Possible Model

The prerequisites for a circular genome to be derived from a linear genome can be listed as follows:

1. A means of producing a circular dsDNA plasmid from linear DNA
2. A mechanism to transfer genes from the linear to the new circular genome
3. Environments that can be colonized because circular genome structure is favored over linear

This could be basically achieved via a ‘‘plasmid transfer’’ model (Fig. 4) with a series of integration events. Processed RNA transcripts (mRNA, rRNA, and tRNA) are the substrate for a reverse transcriptase enzyme which produces an intron-less dsDNA copy which is



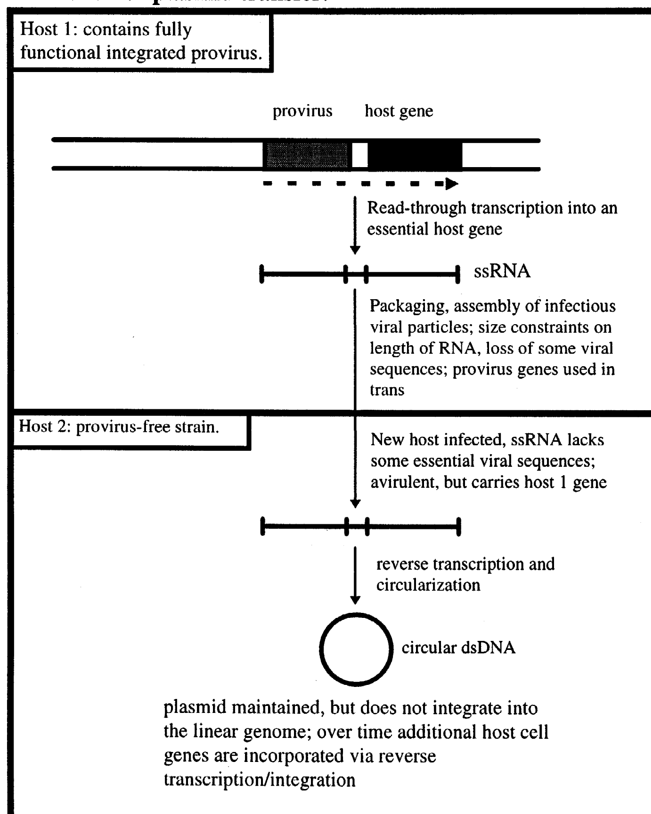


Fig. 4. The plasmid transfer model for the origin of circular genomes. **A** The overall process which is postulated to have occurred. This change in genome structure also resulted in the removal of many aspects of RNA processing from the metabolism of the prokaryote lineage(s) and the loss of intronic sequences. **B** Possible mechanism of plasmid transfer. An infectious agent, similar to a retrovirus, is considered the most likely mechanism by which this change could have been effected. By a process of read-through transcription into a downstream host gene, part of the essential viral genetic information is replaced by a processed host gene. The dsDNA plasmid produced does not integrate into the host genome. The plasmid is maintained as it carries a host gene that is selected for, and the viral reverse transcriptase encoded on the plasmid provides the means for incorporation of processed mRNAs into the new circular chromosome.

integrated into the plasmid. With selection pressures acting upon the organism (*r* selection or thermoreduction—see following sections), this new architecture is favored, ultimately resulting in complete transfer of the genetic information to the circular chromosome. It is possible that the operon control system developed at this time.

Retroviruses may allow an experimental testing of this model because they encoded reverse transcriptase and integrase enzymes and the integrated virus is not excised during the life cycle. Transcription produces a linear single-stranded copy of the RNA which is used in translation to produce the gene products necessary for virion production, and this RNA is also packaged into the virion as the genetic material for the subsequent round of infection. The biology of retroviruses is reviewed in Varmus and Brown (1989). This model requires the produc-

tion of a stable, avirulent, circular version of the retroviral dsDNA that does not become incorporated into the genome of the host but is maintained extrachromosomally as a plasmid (Fig. 4B). The plasmid should have a functional copy of the reverse transcriptase gene. In some instances such circular molecules are produced in vivo in infected cells, though, depending on the system, they may (Panganiban and Temin 1984) or may not (Ellis and Bernstein 1989) be the substrate for integration.

As described in Fig. 4B, the model depicts a read-through error occurring during transcription that results in synthesis of an RNA transcript with a host gene 3′ of the proviral sequences. Packaging then results in the incorporation of an essential host gene (in processed form) into the virus particle, most probably at the expense of an essential viral function, a well-documented phenomenon (Varmus and Brown 1989). The particle then infects a new host, and a circular dsDNA form of the genome is produced which does not become integrated into the host genome but remains extrachromosomal. The precise mechanism of infection is not important; even simple cytoplasmic transfer between host and recipient may be sufficient for infectious agents to spread.

As the host gene would consequently be present in duplicate (one in processed form in the plasmid, and one in the newly infected host) there must be selection that favors either the use of the introduced processed gene over the existing genomic copy (under conditions of *r* selection) or the use of circular dsDNA as a genome architecture per se (such as thermal constraints on the robustness of the genome). These two pressures, *r* selection and adaptation to a thermal environment, are not necessarily mutually exclusive. In fact it is quite likely that they acted concurrently on the emerging prokaryotic lineage(s).

Explanatory Power of Plasmid-Transfer Hypothesis

If current prokaryotic coding genes arose by reverse transcription of mRNA, thus accounting for the absence of introns and the need for a splicing apparatus, then it is likely that operons arose during the same process; genes transcribed into the circular DNA would have an additional advantage if the position in the genome allowed improved coordinate gene expression necessary for the rapid response to various environmental conditions required by an organism under *r* selection. tRNAs would also be reverse transcribed into DNA with the -CCA tail already included, thus bypassing the additional synthetic step of eukaryotes (see Maizels and Weiner 1994).

The selection pressures which produced the advantage for the first circular genome might also have caused the loss of many RNA functions, including that of snoRNA. In addition to the loss described here due to reverse transcription of processed transcripts, additional explanation of this loss is provided by the thermoreduction hypothesis (Forterre 1995a), which is considered next.

Thermophile Life Cycle, the Thermoreduction Hypothesis

Thermophiles (organisms living above about 65°C) are a diverse group of organisms (Cowan 1995)—not one, however is a eukaryote. This is not surprising, given the sensitivity of RNA to increasing rates of hydrolysis above 50°C (Table 1 of Forterre 1995a). In a typical eukaryotic cell it takes 30 min or more for an individual mRNA molecule to be synthesized, capped, spliced, polyadenylated, and transported into the cytoplasm (reviewed in Darnell 1982). Because the half-life of single-stranded RNA is short, about 4 min at 80°C for an RNA molecule with 2,000 nucleotides (Forterre 1995a), mRNA precursors would be rapidly hydrolyzed in a thermophile—unless all the RNA processing steps were removed and protein synthesis was initiated before the mRNA was even fully transcribed—just as in prokaryotes! There is very strong selection for the elimination of ssRNA in thermophiles.

Forterre (1995a, 1996) proposes the ''thermoreduction hypothesis'': Thermophiles are derived from mesophilic ancestors and much of the prokaryotic genome organization results from adaptations to thermophily. Strong advantage accrues to any organism living at high temperature that reduces its processing times for both mRNA and rRNA. As discussed later under *r* and *K* selection, once a prokaryotic genome organization is established, it has other advantages that could be exploited by mesophilic organisms derived from these thermophilic ancestors.

As a result of the many processing steps, and the spatial and temporal separation of transcription and ribosome assembly, the time required to synthesize and assemble a single ribosome is in excess of 1 h in eukaryotes (Girard et al. 1965; Joklik and Becker 1965). At temperatures in which thermophiles thrive, such a molecule would have been hydrolyzed before its processing was finished! In addition, post-transcriptional modification of individual tRNA bases in hyperthermophiles makes these molecules more resistant to degradation at high temperature (Edmonds et al. 1991). Forterre (1995b) suggests that such modifications, and therefore even thermophily itself, were not possible until the advent of tRNA modification enzymes. The small nucleolar RNAs (snorps) involved in ribosome assembly in eukaryotes would also be susceptible to hydrolysis at high temperature, and again are absent from prokaryotes. One of their main functions is specifying methylation of ribose in rRNA, and the function of methylation may be to increase hydrophobicity of the RNA (Kiss-László et al. 1996). It is likely that proteins have taken over the stabilization function in prokaryotes, permitting the loss of snoRNAs.

Table 3. snoRNAs in ribosome maturation and function[a]

	Cleavage of pre-rRNAs	rRNA folding and assembly	Pre-rRNA methylation	Post-assembly scaffold
RNA world	snoRNA	snoRNA	snoRNA	snoRNA
RNP world	RNA + protein	RNA + protein	RNA + protein	Protein
Eukaryotic lineage	RNA + protein	RNA + protein	RNA + protein	Protein
Prokaryotic lineage	Protein*	Protein	Protein	Protein

[a] The hypothetical roles of snoRNAs in the RNA and RNP worlds are suggested from current usage and the general replacement of RNA by protein
[b] The archaeon *Sulfolobus acidocaldarius* has a U3 snoRNA homolog which is essential for rRNA maturation (Potter et al. 1995)

The degradation of RNA at high temperature does not render RNA unusable for thermophiles; clearly, it is possible for RNA to be stabilized at high temperatures otherwise thermophiles would not exist. However, a 3-D structure is expected to be easier to stabilize than single-stranded RNA, because a 3-D structure can be stabilized with additional G-C pairings, methylation of RNA (Forterre 1995a), and/or greater stabilization from proteins. For proteins, only a small number of additional salt bridges, hydrophobic interactions, H-bonds, and increased proline content, together with a reduction in asparagine, are required to increase the relative stability of enzymes (Coolbear et al. 1992; Cowan 1995). Brown et al. (1993) give a similar list for increased stability of the RNA moiety of RNase P: increased number of H-bonds in helices; additional base pairing at base of stem loops; shortened connections between helices; minimization of irregularities (and non–Watson-Crick pairings) in helices. Altogether this minimizes the number of possible alternative foldings. It is the intermediates of rRNA and mRNA (i.e., pre-rRNA and pre-mRNA) that would be more sensitive to hydrolysis at high temperatures; they are less ordered and less stabilized by proteins.

Have prokaryotes been through a thermophilic stage as suggested by Forterre (1995a)? Many groups among eocytes, methanogens, and eubacteria (such as *Aquifex* and *Thermotoga*) are thermophiles. Given our conclusion, based on relics from the RNA world, that the prokaryotic gene organization is derived, it is an attractive idea that a thermophilic stage, even if not extreme thermophily, occurred during the evolution of prokaryotes. Thermoreduction gives a mechanism that explains the following features of prokaryotes:

1. Prokaryotes rapidly process and use their single-stranded messenger and ribosomal RNAs. The mRNAs are translated even before transcription is complete, and prokaryotic rRNA processing appears to begin before the rRNA is completely transcribed (Morrissey and Tollervey 1995).
2. Prokaryotes lack some processing steps entirely. By losing introns from mRNAs and intervening sequences from rRNA (Pace and Burgin 1990) and by the addition of the 3′ terminal CCA into the genomic tRNA sequence, several time-consuming steps of RNA processing have been eliminated. These differences are illustrated in Fig. 2 of Jeffares et al. (1997).
3. Prokaryotes make limited use of RNA as *processors,* such as small nuclear and nucleolar RNPs (see Fig. 3 of Jeffares et al. 1997). Clearly, there is no requirement for a spliceosome once introns have been lost, but it is unclear whether the reduced processing time for mRNA and rRNA or the limited stability of RNA at elevated temperatures would best account for the loss of RNA processors in other roles. The loss of the small nucleolar RNAs is one of the clearest examples of the replacement of RNA by protein (Table 3). Small nucleolar RNAs are implicated in a number of essential aspects of the biogenesis of ribosomes, including pre-rRNA cleavage, methylation, and ribosome assembly. We suggest that an ancient role for snoRNAs may have been as a scaffold maintaining the active structure of the mature ribosome, particularly since some snoRNAs (U32 and U36) have regions of complementarity to both the 18S and 28S rRNAs (Nicoloso et al. 1996). Intermolecular linking is not required per se for methylation of rRNA—only complementarity with rRNA is required (Kiss-László et al. 1996). This point is discussed in greater detail in Jeffares et al. (1997). We expect that the scaffold snoRNA function was the first to be supplanted by proteins, as shown in Table 3, which can be envisaged as a time series illustrating the various roles that snoRNAs have performed as well as their gradual replacement by proteins.
4. The unusual ‘‘mischarging’’ of $tRNA^{Gln}$ with glutamate and the subsequent conversion of the bound glutamate to glutamine (Rogers and Söll 1995) occurs in archaea and gram-positive eubacteria. Glutamine, and amides generally, are unstable at high temperatures (Greenstein and Winitz 1961), and so free glutamine would only exist in the cell of thermophiles at extremely low concentrations. This would limit the rate of protein synthesis unless an alternative source is available. A solution would be for thermophiles to charge $tRNA^{Gln}$ with glutamic acid and transaminate it immediately before incorporating glutamine into protein. Gram-negative bacteria would later have re-established the original direct pathway of incorporation of glutamine either by horizontal gene transfer

or by duplication and subsequent functional divergence of an existing tRNA synthetase gene, perhaps $tRNA^{Glu}$ synthetase (Lamour et al. 1994; Gagnon et al. 1996). Thus we consider mischarging as a biochemical adaptation to extreme conditions, not an ancestral feature.
5. Closed circular DNA is used as the genetic material in prokaryotes. Linear genomes with very simple telomeres are present in some bacterial cells (Hinnebusch and Tilly 1993), so there is no absolute requirement for a circular chromosome in prokaryotes. Since the melting temperature of closed circular DNA at physiological salt concentrations, irrespective of supercoiling, is significantly higher than that of an equivalent linear DNA (Vinograd et al. 1968; Sen et al. 1992; Marguet and Forterre 1994), a circular genome organization in both prokaryotic lineages may also be an adaptation to thermophily (see next section). Furthermore, the enzyme reverse gyrase that increases the linking number of covalently closed circular dsDNA and appears essential for life at high temperatures is probably the result of a gene fusion event involving ancient helicase and topoisomerase I genes (Forterre et al. 1995). Consequently it must have arisen later in evolution, apparently in the process of adaptation to thermophily. The introduction of circular genomes also releases cells from the need for the RNP telomerase, which may have a (potentially heat sensitive) RNA catalytic site (Gilley et al. 1995; McEachern and Blackburn 1995; Gilley and Blackburn 1996).

Several curious features of RNase P enzymes appear to be related to thermophily. Thus far only prokaryotic RNase P RNAs have shown catalytic function in the absence of the C5 protein subunit—eukaryotic M1 RNAs appear to require the protein subunit for activity in vitro (Forster and Altman 1990, and references therein). The optimum activity for the RNase P RNAs from *Thermus aquaticus, Thermotoga maritima,* and *E. coli* is achieved with 3 M ammonium chloride at a temperature of 60°C, which appears peculiar for an organism that had never had a thermophilic ancestor. Optimum reaction temperatures for RNase P RNA from the two thermophiles is only 5–10°C higher than that found for *E. coli* (Brown et al. 1993). With the ribosome we may expect similar features, both within and between proteins and RNAs, that increase stability at high temperature. Given this, it seems imperative to examine the heat stability of the ribosomal rRNAs in the same manner as Brown et al. (1993), to search for any further evidence of an early thermophilic stage in the evolution of prokaryotic life.

r Selection in Prokaryotes

The second line of reasoning in favor of prokaryotic genomes being derived is *r* selection in early prokaryotes. The concepts of *r*- and *K*-selection (Pianka 1970) are important in ecological and life history research. The terms come from the equation for rate of population growth, *R.*

$$R = dN/dt = rN\,(1 - N/K)$$

where *r* is the maximum intrinsic rate of increase for the population, *N* the number of organisms, and *K* the carrying capacity. *r*-selected organisms show high reproductive rates, small size, short life cycles, and are often found in unpredictable environments. (They respond rapidly to an increased food supply.) In contrast, *K*-selected organisms show the reverse of these properties—longer life cycles, larger size, relatively more stable population sizes, and are found in more stable environments (Pianka 1970). In general, prokaryotes are much more *r*-selected than eukaryotes in that generation time is shorter, genome size is smaller, and population size is larger (and subject to dramatic changes in size). Eukaryotes are variable in the *r–K* selection continuum, with popular organisms for genetic studies (yeast, *Caenorhabditis, Drosophila,* mice, and *Arabidopsis*) being more highly *r*-selected eukaryotes.

Selection for small genome size and faster gene expression in prokaryotes favors the replacement of large RNP complexes, the loss of nonessential genetic material (such as introns and the spliceosome), and a more efficient use of genome space; the *E. coli* genome is almost 100% coding DNA, whereas eukaryotes may range from 70% (yeast) to less than 1% coding DNA (the lungfish *Protopterus,* and *Fritillaria,* a flowering plant) (Maynard Smith and Szathmáry 1995). Correlations between developmental rate and genome size are evident within certain taxonomic groups, supporting the view that ''junk-DNA'' accumulates until it becomes a burden to maintain (Pagel and Johnstone 1992). The important factor for *r* selection is the *ability* to reproduce quickly under appropriate conditions. A single center of replication also favors reduced genome size in *r*-selected organisms because the time required to copy the entire genome becomes longer than the cell division cycle. A second or even third round of replication may be started before the first is completed, and as this continues, genes will occur in different dosages. Separation of the different genomes would increase complications. Increasingly, larger genome size will eventually limit the rate of reproduction. Under conditions of rapid growth and with a single center of replication, there is strong selection to limit genome size.

Summary and Discussion

We have drawn upon the molecular fossil record which was initially used to build a model for the last riboorganism (Jeffares et al. 1997). Taking this model, and

with due consideration to the genetics and metabolism of extant organisms, we have described a path from the RNA world, via the breakthrough organism that developed protein synthesis, to the last universal common ancestor. This model has enabled us to examine the nature of the LUCA and describe a testable model that explains the origin of the prokaryotic genome architecture and its lack of RNA-dependent RNA metabolism.

One of the most interesting corollaries of the RNA world model (Jeffares et al. 1997) is that it allows formation of a detailed model for the origin of protein synthesis; a high-accuracy RNA polymerase that added triplets of nucleotides would have most of the steps required for protein synthesis by ribosomes. Several versions of the model are possible; the most detailed ones have the triplet code established before protein synthesis. Additionally, the introns-first hypothesis proposed here provides a concrete and testable model to explain the genomic origin of the first mRNAs and the nature of the first genetically coded proteins (chaperone-like). Furthermore, this model allows us to suggest examples of such proteins by their physical contiguity with confirmed fossil RNAs.

The combined effects of thermoreduction (Forterre 1995a, 1996) and *r* selection account for differences between the genomes of prokaryotes and eukaryotes and for the loss of RNPs in prokaryotes. An important conclusion of our study is that the reduced number of RNPs in prokaryotes indicates that they have been derived from an ancestor with a eukaryote-like genome organization; this approach does not provide any information about the cytological organization of this organism. The structure of prokaryotic genomes thus appears to be derived, a eukaryote-type genome being more feasible as a primitive genetic system. The model includes both a mechanism (plasmid transfer via reverse transcriptase) and selective advantages—namely, faster response times, coordinate control of groups of genes, and/or survival at higher temperatures (thermophily). We therefore consider it straightforward for a prokaryotic genome organization to develop in a normal evolutionary process.

The simplest explanation is that the circular genome arose once, but we cannot exclude that it arose separately in eubacteria and archaea (Forterre 1996). The presence of reverse gyrase in both the archaeal and eubacterial hyperthermophiles (reviewed in Forterre 1996) could allow tests of prokaryotic monophyly. To help resolve this issue is necessary to examine more widely the distribution of this enzyme in thermophiles and to search for this enzyme in mesophiles related to thermophilic organisms. (For example, members of the *Deinococcus-Thermus* group—Van den Eynde et al. 1990; Eisen 1995). Furthermore it would be pertinent to investigate the possibility that this gene has been subjected to a horizontal transfer event.

An interesting consequence of the thermostability of enzymes is that because additional H-bonds, salt bridges, etc., are required to stabilize 3-D structures at high temperatures (or under adverse conditions generally), additional constraints are placed upon amino acids or nucleotides regarding freedom to change during evolution. These additional constraints would lead to lower rates of molecular evolution in archaea (Dickerson 1971) and indeed lower rates have been observed in archaea (Lake 1987). The instability of RNA at higher temperatures also has implications for the origin of life. Several authors have pointed out that a thermophilic origin of life is inconsistent with an RNA world (Joyce 1988; Miller and Bada 1988; Forterre 1992).

The arguments presented in this paper represent an alternative to using sequence data for inferring the earliest features of living systems. Many authors have used sequence information for determining deep divergences (Gupta and Singh 1994; Lazcano and Miller 1994; Gogarten et al. 1989; Iwabe et al. 1989; Baldauf et al. 1996). We are not yet convinced that current methods of reconstructing trees from sequences are adequate for such ancient divergences. Rates of molecular evolution may vary for many reasons, including generation time effects (Li et al. 1996), but a realistic average rate of evolution for neutral sites is about 0.5×10^{-6} changes per site per year. At this rate, two sequences sharing a common ancestor 3 billion years ago will differ by an average of about 30 changes at every site that is unconstrained (free to vary) over the whole time period. It is far beyond the ability of current tree-inference methods to recover such trees accurately with finite-length sequences (Charleston et al. 1994), especially as other processes such as varying nucleotide composition and differences in which sites cannot vary mislead methods for inferring trees (Lockhart et al. 1994, 1996). The disagreement over the position of Microsporidia (Kamaishi et al. 1996; Keeling and Doolittle 1996) exemplifies the problems in using protein or rRNA sequence data to resolve deep divergences. In reality, most sites will probably not be free to vary over the full time period because selection is operating. This is the covarion model (Fitch and Ayala 1994) which offers an explanation as to how *sometimes* a tree reconstruction method could correctly recover such an ancient divergence. However, the models currently in use do not predict good performance for ancient divergences and so, until better models for sequence data are available, the simpler methods used here are the best (Penny et al. 1994; Lockhart et al. 1996).

The early stromatolite fossils appear not to support the idea that prokaryotes arose first. However, the evidence that the oldest stromatolites are cyanobacterial (Walsh 1992) is neither direct nor conclusive (Lowe 1994; Grotzinger and Rothman 1996), and under some theories of the evolution of photosynthesis (Larkum 1991), cyanobacteria are a relatively recent photosynthetic group that has replaced earlier forms using chlorophyll *b* and/or

c as accessory pigments. Nor can the stromatolite fossils be compared genetically or biochemically to extant cyanobacteria. Morphological similarities between fossils in stromatolites and modern-day cyanobacteria do not necessarily link them from a taxonomic point of view. Finally, a thermophilic prokaryote last universal common ancestor is incompatible with the RNA world theory (Forterre 1995b; Forterre et al. 1995).

Hypotheses have many useful roles in science including making predictions, solving puzzles, showing connections between apparently unrelated phenomena, and generating new ideas for testing. In the present case the primary aim was to find whether a logically consistent model, based on a wide range of molecular and theoretical information from extant life, was possible both for the RNA world (Jeffares et al. 1997) and for the path to the last universal common ancestor. In these aspects we feel that the model succeeds. Although the model as a whole will be difficult to test directly, it is certainly helpful if it leads to new studies. Some possibilities are: the presence/absence of RNase MRP in Microsporidia, Giardia and Entamoeba; whether the RNA of MRP is large like eukaryotic RNase P, or small as in eubacteria; checking for the presence of telomerase in plants; the involvement of telomerase RNA in the catalytic center of telomerase as evidence for linear genomes (and telomeres) as ancestral. A test for the ''introns-first'' theory would examine whether the de novo evolution of a low-complexity protein from nonfunctional mRNA containing an intronic snoRNA (such as that described by Tycowski et al. 1996) is possible. This would provide an experimental means of examining whether function may evolve de novo in genomically discrete, noncontiguous fragments, which are ligated at the messenger level.

An examination of the phylogenetic distribution of glutamine tRNA synthetase would be of interest, given that mischarging of $tRNA^{Gln}$ is likely to be an adaptation to thermophily. We would not expect to find this phenomenon in eukaryotes, but we would expect it to occur in all thermophiles. Current genome sequencing efforts may very soon make available the tools to carry out extensive studies into this.

The plasmid transfer hypothesis could be tested experimentally through the use of retroviral genetics to see if there is any selection pressure (such as exposure of the host to elevated temperatures or conditions favoring *r* selection) that would repeatedly result in the transfer of host genes to such a stable, avirulent retroviral plasmid. Possibly, a eukaryote that is tolerant of moderate temperature increases, such as *Thermophilus thermotoga,* may make an appropriate host for such experiments. Additionally, if a successful assay system could be developed, it would potentially be possible to examine the effects of other extreme environmental pressures, such as high or low pH, limited nutrients, or high salt conditions on such a system. This may then provide clues as to the feasibility of the circular genome being derived. Using the RNA world as an outgroup to root the Tree of Life is thus a potentially interesting alternative to current hypotheses.

Acknowledgments. We thank Patrick Forterre and an anonymous referee for valuable and constructive comments on the manuscript.

References

Baldauf SL, Palmer JD, Doolittle WF (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

Baumann P, Qureshi SA, Jackson SP (1995) Transcription: new insights from studies on archaea. Trends Genet 11:279–283

Brown JW, Haas ES, Pace NR (1993) Characterisation of ribonuclease P RNAs from thermophilic bacteria. Nucleic Acid Res 21:671–679

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb J-F, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Weidman JF, Fuhrmann JL, Nguyen D, Utterback TR, Kelley JM, Peterson JD, Sadow PW, Hanna MC, Cotton MD, Roberts KM, Hurst MA, Kaine BP, Borodovsky M, Klenk H-P, Fraser CM, Smith HO, Woese CR, Venter JC (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii.* Science 273:1058–1072

Carlile MJ (1980) From prokaryote to eukaryote: gains and losses. Symp Gen Microbiol 30:1–40

Carlile MJ (1982) Prokaryotes and eukaryotes: strategies and successes. Trends Biochem Sci 7:128–130

Charleston MA, Hendy MD, Penny D (1994) The effects of sequence length, tree topology and number of taxa on the performance of phylogenetic methods. J Comput Biol 1:133–151

Collins K, Greider CW (1993) *Tetrahymena* telomerase catalyzes nucleolytic cleavage and nonprocessive elongation. Genes Dev 7: 1364–1376

Collins K, Kobayashi R, Greider CW (1995) Purification of *Tetrahymena* telomerase and cloning of genes encoding the two protein components of the enzyme. Cell 81:677–686

Coolbear T, Daniel RM, Morgan HW (1992) The enzymes from extreme thermophiles: bacterial sources, thermostabilities and industrial relevance. Adv Biochem Eng Biotechnol 45:57–98

Cooperman BS, Romero DP, Traut RR (1995) Histidine 229 in protein L2 is apparently essential for peptidyl transferase activity. In: Program Frontiers in Translation, Canada, May 1995

Cowan DA (1995) Protein stability at high temperatures. Essays Biochem 29:193–207

Crick FHC (1968) The origin of the genetic code. J Mol Biol 38:367–379

Darlington CD (1958) The evolution of genetic systems. 2nd ed. Cambridge University Press, Cambridge

Darnell JE (1981) Do features of present-day eukaryotic organisms reflect ancient sequence arrangements? In: Scudder GGE, Reveal JL (eds) Evolution today, Proceedings of the Second International Congress of Systematic and Evolutionary Biology, pp 207–213

Darnell JE (1982) Variety in the level of gene control in eukaryotic cells. Nature 297:365–371

Dickerson RE (1971) The structure of cytochrome c and the rates of molecular evolution. J Mol Evol 1:26–45

Doolittle RF (1995) Of Archae and Eo: what's in a name? Proc Natl Acad Sci USA 92:2421–2423

Doudna JA, Szostak JW (1989) RNA-catalysed synthesis of complementary-strand RNA. Nature 339:519–522

Draper DE (1995) Protein-RNA recognition. Annu Rev Biochem 64: 593–620

Edmonds CG, Crain PF, Gupta R, Hashizume T, Hocort CH, Kowalak JA, Pomerantz SC, Stetter KO, McCloskey JA (1991) Posttranscriptional modification of tRNA in thermophilic archaea (archaebacteria). J Bacteriol 173:3138–3148

Eigen M (1992) Steps toward life: a perspective on evolution. Oxford University Press, Oxford

Eigen M (1993) The origin of genetic information: viruses as models. Gene 135:37–47

Eigen M, Schuster P (1978) The hypercycle: a principle of natural self-organization, Part C the realistic hypercycle. Naturwissenschaften 65:341–369

Eigen M, Winkler-Oswatitsch R (1981) Transfer-RNA, an early gene? Naturwissenschaften 68:282–292

Eisen JA (1995) The RecA protein as a model module for molecular systematic studies of bacteria: comparison of trees of RecAs and 16s rRNAs from the same species. J Mol Evol 41:1105–1123

Ekland EH, Bartel DP (1996) RNA-catalysed RNA polymerisation using nucleoside triphosphates. Nature 382:373–376

Ellis J, Bernstein A (1989) Retrovirus vectors containing an internal attachment site: evidence that circles are not intermediates to murine retrovirus integration. J Virol 63:2844–2846

Fangman WL, Brewer BJ (1991) Activation of replication origins within yeast chromosomes. Annu Rev Cell Biol 7:375–402

Fitch WM, Ayala FJ (1994) The superoxide dismutase molecular clock revisited. Proc Natl Acad Sci USA 91:6802–6807

Forster AC, Altman S (1990) Similar cage-shaped structures for the RNA components of all ribonuclease P and ribonuclease MRP enzymes. Cell 62:407–409

Forterre P (1992) New hypotheses about the origins of viruses, prokaryotes and eukaryotes. In: Trân Thanh Vân IK, Mounolo JC, Schneider I, McKay C (eds) Frontiers of life. Éditions Frontières, Gif-sur-Yvette, France pp 221–234

Forterre P (1995a) Thermoreduction, a hypothesis for the origin of prokaryotes. C R Acad Sci Paris Life Sci 318:415–422

Forterre P (1995b) Looking for the most ''primitive'' organism(s) on Earth today: the state of the art. Planet Space Sci 43:167–177

Forterre P (1996) A hot topic: the origin of hyperthermophiles. Cell 85:789–792

Forterre P, Confalonieri F, Charbonnier F, Duguet M (1995) Speculations on the origin of life and thermophily: review of available information on reverse gyrase suggests that hyperthermophilic procaryotes are not so primitive. Orig Life Evol Biosph 25:235–249

Gagnon Y, Lacoste L, Champagne N, Lapointe J (1996) Widespread use of the Glu-tRNA^{Gln} transamidation pathway among bacteria. J Biol Chem 271:14856–14863

Gilbert W, Glynias M (1993) On the ancient nature of introns. Gene 135:137–144

Gilley D, Blackburn EH (1996) Specific RNA residue interactions required for enzymatic functions of *Tetrahymena* telomerase. Mol Cell Biol 16(1):66–75

Gilley D, Lee MS, Blackburn EH (1995) Altering specific telomerase RNA template residues affects active site function. Genes Dev 9:2214–2226

Girard M, Latham H, Penman S, Darnell JE (1965) Entrance of newly formed messenger RNA and ribosomes into HeLa cell cytoplasm. J Mol Biol 11:187–201

Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF, Poole RJ, Date T, Oshima T, Konishi J, Denda K, Yoshida M (1989) Evolution of the vacuolar H+-ATPase: implications for the origin of eukaryotes. Proc Natl Acad Sci USA 86: 6661–6665

Gordon KHJ (1995) Were RNA replication and translation directly coupled in the RNA (+protein?) world? J Theor Biol 173:179–193

Greenstein JP, Winitz M (1961) Ch 25. Glutamic acid and glutamine. In: Chemistry of the amino acids, vol 3. John Wiley, New York, pp 1929–1954

Grotzinger JP, Rothman DH (1996) An abiotic model for stromatolite morphogenesis. Nature 383:423–425

Gupta RS, Singh B (1994) Phylogenetic analysis of 70kD heat shock protein sequences suggests a chimaeric origin for the eukaryotic nucleus. Curr Biol 4:1104–1114

Hartl FU (1996) Molecular chaperones in cellular protein folding. Nature 381:571–580

Härtlein M, Cusack S (1995) Structure, function and evolution of seryl-tRNA synthetases: implications for the evolution of aminoacyl-tRNA synthetases and the genetic code. J Mol Evol 40:519–530

Hinnebusch J, Tilly K (1993) Linear plasmids and chromosomes in bacteria. Mol Microbiol 10:917–922

Illangasekare M, Sanchez G, Nickles T, Yarus M (1995) Aminoacyl-RNA synthesis catalyzed by an RNA. Science 267:643–647

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

Jeffares DC, Poole AM, Penny D (1995) Pre-rRNA processing and the path from the RNA world. Trends Biochem Sci 20:298–299

Jeffares DC, Poole AM, Penny D (1997) Relics from the RNA world. J Mol Evol 45:18–36

Joklik WK, Becker Y (1965) Studies on the genesis of polyribosomes. I. Origin and significance of the subribosomal particles. J Mol Biol 13:496–510

Joyce GF (1988) Hydrothermal vents too hot? Nature 334:564

Jurka J, Smith TF (1987) β-Turn-driven early evolution: the genetic code and biosynthetic pathways. J Mol Evol 25:15–19

Kamaishi T, Hashimoto T, Nakamura Y, Nakamura F, Murata S, Okada N, Okamoto K, Shimizu M, Hasegawa M (1996) Protein phylogeny of translation elongation factor EF-1α suggests microsporidians are extremely ancient eukaryotes. J Mol Evol 42:257–263

Keeling PJ, Doolittle WF (1996) Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. Mol Biol Evol 13:1297–1305

Keese PK, Gibbs A (1992) Origins of genes: 'big bang' or continuous creation? Proc Natl Acad Sci USA 89:9156–9160

Kiss T, Filipowicz W (1993) Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene RCC1. EMBO J 12:2913–2920

Kiss-László Z, Henry Y, Bachellerie J-P, Caizergues-Ferrer M, Kiss T (1996) Site-specific ribose methylation of preribosomal RNA; a novel function for small nucleolar RNAs. Cell 85:1077–1088

Koch AL (1984) Evolution vs the number of gene copies per primitive cell. J Mol Evol 20:71–76

Lake JA (1987) Prokaryotes and archaebacteria are not monophyletic: rate invariant analysis of rRNA genes indicates that eukaryotes and eocytes form a monophyletic taxon. Cold Spring Harb Symp Quant Biol 52:839–846

Lamour V, Quevillon S, Diriong S, N'Guyen VC, Lipinski M, Mirande M (1994) Evolution of the Glx-tRNA synthetase family: the glutaminyl enzyme as a case of horizontal gene transfer. Proc Natl Acad Sci USA 91:8670–8674

Larkum AWD (1991) The evolution of chlorophylls. In: Scheer H (ed) Chlorophylls. CRC Press, Boca Raton, FL, pp 3–30

Lazcano A, Miller SL (1994) How long did it take for life to begin and evolve to cyanobacteria? J Mol Evol 39:546–554

Leverette RD, Andrews MT, Maxwell ES (1992) Mouse U14 snRNA is a processed intron of the cognate hsc70 heat shock pre-messenger RNA. Cell 71:1215–1221

Li W-H, Ellsworth DL, Krushkal J, Chang BH-J, Hewett-Emmet D (1996) Rates of substitution in primates and rodents and the generation-time effect hypothesis. Mol Phyl Evol 5:182–187

Liu J, Maxwell ES (1990) Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene. Nucleic Acids Res 18:6565–6571

Lockhart PJ, Steel MA, Hendy MD, Penny D (1994) Recovering evo-

lutionary trees under a more realistic model of sequence evolution. Mol Biol Evol 11:605–612

Lockhart PJ, Larkum AWD, Steel MA, Waddell PJ, Penny D (1996) Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. Proc Natl Acad Sci USA 93: 1930–1934

Loidl J (1991) Coming to grips with a complex matter. Chromosoma 100:289–292

Lowe DR (1994) Abiological origin of described stromatolites older than 3.2 Ga. Geology 22:387–390

Lund PA (1995) The roles of molecular chaperones *in vivo.* Essays Biochem 29:113–123

Maizels N, Weiner AM (1987) Peptide-specific ribosomes, genomic tags, and the origin of the genetic code. Cold Spring Harb Symp Quant Biol 52:743–749

Maizels N, Weiner AM (1993) The genomic tag hypothesis: modern viruses as molecular fossils of ancient strategies for genomic replication. In: Gesteland RF, Atkins JF (eds) The RNA world. Cold Spring Harbor Laboratory Press, New York, pp 577–602

Maizels N, Weiner AM (1994) Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. Proc Natl Acad Sci USA 91:6729–6734

Marguet E, Forterre P (1994) DNA stability at temperatures typical for thermophiles. Nucleic Acids Res 22:1681–1686

Maxwell ES, Fournier MJ (1995) The small nucleolar RNAs. Annu Rev Biochem 35:897–934

Maynard Smith J, Szathmáry E (1995) The major transitions in evolution. W H Freeman, Oxford

McEachern MJ, Blackburn EH (1995) Runaway telomere elongation caused by telomere RNA gene mutations. Nature 376:403–409

Miller SL, Bada JL (1988) Submarine hot springs and the origin of life. Nature 334:609–611

Morrissey JP, Tollervey D (1995) Birth of the snoRNPs: the evolution of RNase MRP and the eukaryotic pre-rRNA-processing system. Trends Biochem Sci 20:78–82

Nagel GM, Doolittle RF (1995) Phylogenetic analysis of the aminoacyl-tRNA synthetases. J Mol Evol 40:487–498

Nicoloso M, Qu LH, Michot B, Bachellerie J-P (1996) Intron-encoded, antisense small nucleolar RNAs—the characterization of nine novel species points to their direct role as guides for the 2-O-methylation of rRNAs. J Mol Biol 260:178–195

Orgel LE (1986) RNA catalysis and the origins of life. J Theor Biol 123:127–149

Pace NR, Burgin AB (1990) Processing and evolution of the rRNAs. In: Hill WE, Dahlberg A, Garrett RA, Moore PB, Schlessinger D, Warner JR (eds) The ribosome: structure, function, evolution. American Society for Microbiology, Washington, pp 417–425

Pagel M, Johnstone RA (1992) Variation across species in the size of the nuclear genome supports the junk-DNA explanation of the C-value paradox. Proc R Soc Lond Biol 249:119–124

Palmer JD, Logsdon JM (1991) The recent origin of introns. Curr Opin Genet Dev 1:470–477

Panganiban AT, Temin HM (1984) Circles with two tandem LTRs are precursors to integrated retrovirus DNA. Cell 36:673–679

Penny D (1985) The evolution of meiosis and sexual reproduction. Biol J Linn Soc 25:209–220

Penny D, Lockhart PJ, Steel MA, Hendy MD (1994) The role of models in reconstructing evolutionary trees. In: Siebert D (ed) Models in phylogeny. Oxford University Press, Oxford, pp 211–230

Pianka ER (1970) On *r*- and *K*-selection. Am Nat 104:592–597

Piccirilli JA, Krauch T, Moroney SE, Benner SA (1990) Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. Nature 343:33–37

Potter S, Durovic P, Dennis PP (1995) Ribosomal RNA precursor processing by a eukaryotic U3 small nucleolar RNA-like molecule in an archaeon. Science 268:1056–1060

Purohit P, Stern S (1994) Interactions of a small RNA with antibiotic and RNA ligands of the 30S subunit. Nature 370:659–662

Reanney DC (1974) On the origin of prokaryotes. J Theor Biol 48: 243–251

Reanney DC (1979) RNA splicing and polynucleotide evolution. Nature 277:598–600

Reanney DC (1982) The evolution of RNA viruses. Annu Rev Microb 36:47–73

Reanney DC (1987) Genetic error and genome design. Cold Spring Harb Symp Quant Biol 52:751–757

Rogers KC, Söll D (1995) Divergence of glutamate and glutamine aminoacylation pathways: providing the evolutionary rationale for mischarging. J Mol Evol 40:476–481

Schimmel P, de Pouplana LR (1995) Transfer RNA: from minihelix to genetic code. Cell 81:983–986

Sen S, Lahiri A, Majumdar R (1992) Melting characteristics of highly supercoiled DNA. Biophys Chem 42:229–234

Séraphin B (1993) How many intronic snRNAs? Trends Biochem Sci 18:330–331

Sharp PA (1985) On the origin of RNA splicing and introns. Cell 42:397–400

Steitz JA, Tycowski KT (1995) Small RNA chaperones for ribosome biogenesis. Science 270:1626–1627

Switzer C, Moroney SE, Benner SA (1989) Enzymatic incorporation of a new base pair into DNA and RNA. J Am Chem Soc 111:8322–8323

Szewczak AA, Moore PB (1995) The sarcin/ricin loop, a modular RNA. J Mol Biol 247:81–98

Szybalski W, Szybalski EH (1974) Visualisation of the evolution of viral genomes. In: Kurstak E, Maramorosch K (eds) Viruses, evolution and cancer: basic considerations. Academic Press, New York, pp 563–582

Taylor FJR, Coates D (1989) The code within the codons. Biosystems 22:177–187

Tollervey D (1996) Small nucleolar RNAs guide ribosomal RNA methylation. Nature 273:1056–1057

Tycowski KT, Shu M-D, Steitz JA (1993) A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. Gene Dev 7:1176–1190

Tycowski KT, Shu M-D, Steitz JA (1996) A mammalian gene with introns instead of exons generating stable RNA products. Nature 379:464–466

Van den Eynde H, Van de Peer Y, Vandenabeele H, Van Bogaert M, De Wachter R (1990) 5S rRNA sequences of myxobacteria and radioresistant bacteria and implications for eubacterial evolution. Int J Syst Bacteriol 40:399–404

Varmus H, Brown P (1989) Retroviruses. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington, DC, pp 53–108

Vinograd J, Lebowitz J, Watson R (1968) Early and late helix-coil transitions in closed circular DNA. The number of superhelical turns in polyoma DNA. J Mol Biol 33:173–197

Walsh MM (1992) Microfossils and possible microfossils from the Early Archean Onverwacht Group, Barberton Mountain Land, South Africa. Precambrian Res 54:271–293

Weiss R, Cherry J (1993) Speculations on the origin of ribosomal translation. In: Gesteland RF, Atkins JF (eds) The RNA world. Cold Spring Harbor Laboratory Press, New York, pp 71–89

Wetzel R (1995) Evolution of the aminoacyl-tRNA synthetases and the origin of the genetic code. J Mol Evol 40:545–550

Wool IG (1996) Extraribosomal functions of ribosomal proteins. Trends Biochem Sci 21:164–165

Wootton JC (1994) Sequences with ‘unusual’ amino acid compositions. Curr Opin Struct Biol 4:413–421