HYPOTHESIS

Introns and the origin of nucleus–cytosol compartmentalization

William Martin[1] & Eugene V. Koonin[2]

The origin of the eukaryotic nucleus marked a seminal evolutionary transition. We propose that the nuclear envelope's incipient function was to allow mRNA splicing, which is slow, to go to completion so that translation, which is fast, would occur only on mRNA with intact reading frames. The rapid, fortuitous spread of introns following the origin of mitochondria is adduced as the selective pressure that forged nucleus–cytosol compartmentalization.

The discovery of introns had a broad effect on thoughts about early evolution. Walter Gilbert coined the terms introns and exons while predicting that their utility in gene evolution might lie in exon shuffling and alternative splicing[1]. Ford Doolittle suggested that the ancestral state of gene structure might be 'split' and that eukaryotes might have preserved, in the guise of exons, relics from the primordial assembly of genes via recombination in introns[2]. James Darnell submitted similar ideas at a time when the issue in early evolution was how to generate long coding sequences from scratch[3]. Exonic modules provided an answer, and the introns-early view was born: the intron-bearing eukaryotic lineage was seen as an independent primordial line of descent whose genealogy went straight back to the origin of genes, as evidenced by their introns, and whose descendants included the host that acquired the mitochondrion[4]. Because introns-early proposed a modular assembly of genes[5], it readily accommodated the discovery of catalytic RNA[6] and the 'RNA world' concept[7]. But it also predicted exon boundaries in genes to correlate with domain boundaries in proteins[8], and hence was abandoned when that prediction failed[9]. Yet exon shuffling and alternative splicing[1] remain as paradigms of eukaryotic gene evolution, and the effect of introns-early on theories about early evolution persists: the notions that eukaryotic introns might be direct holdovers from the origin of genes and that prokaryotic genomes were once full of introns, but became streamlined, still enjoy some currency[10–12].

Throughout all that, the origins of mRNA introns themselves remained an issue. Cech[13] proposed that eukaryotic introns and their cognate spliceosomal small nuclear RNAs (snRNAs) originated from disarticulate group II introns. Cavalier-Smith[14] extended that idea by suggesting that they evolved specifically from group II introns that invaded the ancestrally intron-less eukaryotic genome through the mitochondrial endosymbiont, thereby generating two predictions. First, group II introns should be found among free-living α-proteobacteria, the ancestors of mitochondria. Second, eukaryotic lineages that were then suspected to primitively lack mitochondria, such as *Giardia intestinalis* (syn. *G. lamblia*), should lack introns.

The first prediction was borne out directly[15], supporting the idea that introns could originate from mitochondria[14,16]. The mobility of group II introns in contemporary eubacteria[17] and their prevalence in α-proteobacteria[18] are still fully consistent with that view. The second prediction turned out to be wrong, but for an unexpected reason: *Giardia* has introns after all[19], but it also has mitochondria[20]. In fact, all of the eukaryotes that had been predicted to lack introns and mitochondria[14] have both[19–24]. So the suggestion that introns and the spliceosome arose in the wake of mitochondrial origin could still be right[14].

Here we revisit the possible evolutionary significance of introns in light of mitochondrial ubiquity. We propose that the spread of group II introns and their mutational decay into spliceosomal introns created a strong selective pressure to exclude ribosomes from the vicinity of the chromosomes—thus breaking the prokaryotic paradigm of co-transcriptional translation and forcing nucleus–cytosol compartmentalization, which allowed translation to occur on properly matured mRNAs only.

New consensus, new observations

A current consensus on introns would be that prokaryotes do indeed have group II introns but that they never had spliceosomes[18]; hence, streamlining in the original sense (that is, loss of spliceosomal introns) never occurred in prokaryotes, although it did occur in some eukaryotes such as yeast or microsporidia[25,26]. An expansion of that consensus would be that spliceosomes and spliceosomal introns are universal among eukaryotes[27], that group II introns originating from the mitochondrion are indeed the most likely precursors of eukaryotic mRNA introns and spliceosomal snRNAs[16,18,28], and that many—conceivably most—eukaryotic introns are as old as eukaryotes themselves[25,26]. More recent are the insights that there is virtually no evolutionary grade detectable in the origin of the spliceosome, which apparently was present in its (almost) fully fledged state in the common ancestor of eukaryotic lineages studied so far[27], and that the suspected source of introns—mitochondria, including their anaerobic forms, hydrogenosomes and mitosomes[20,23,24]—was also present in the common ancestor of contemporary eukaryotes (the only ones whose origin or attributes require explanation).

This suggests that intron origin and spread occurred within a narrow window of evolutionary time: subsequent to the origin of the mitochondrion, but before the diversification of the major eukaryotic lineages. This, in turn, indicates the existence of a turbulent phase of genome evolution in the wake of mitochondrial origin, during which group II introns invaded the host's chromosomes, spread as transposable elements into hundreds—perhaps thousands—of positions that have been conserved to the present, and fragmented into both mRNA introns and snRNA constituents of the spliceosome. Why is this noteworthy?

The organisms that have mitochondria have introns, and a causal link is suspected. But the organisms that have mitochondria and introns also have a cell nucleus, and there are now reasons to suspect a

[1]Institute of Botany III, University of Düsseldorf, D-40225 Düsseldorf, Germany. [2]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA.

causal link here as well. The evolutionary relationships of proteins specific to the nuclear envelope and nuclear pore complex reveal that this sizable protein set is a mix of proteins and domains of archaebacterial and eubacterial origins, along with some eukaryotic innovations, suggesting that the nucleus arose in a cell that already contained a mitochondrial endosymbiont[29]. Furthermore, gene duplication patterns suggest that the endoplasmic reticulum arose before the nucleus[29], as it occurs in the modern cell cycle[30]. Additionally, evolutionary proteomics show that the nucleolus also contains proteins of both archaebacterial and eubacterial origins, suggesting that it too arose in a mitochondriate cell[31]. None of the foregoing interpretations needs be correct. However, biologists have traditionally addressed the origin of the nucleus independently of mitochondrial origins[30], which need not be correct either. Hence, it is prudent to explore alternative avenues of investigation.

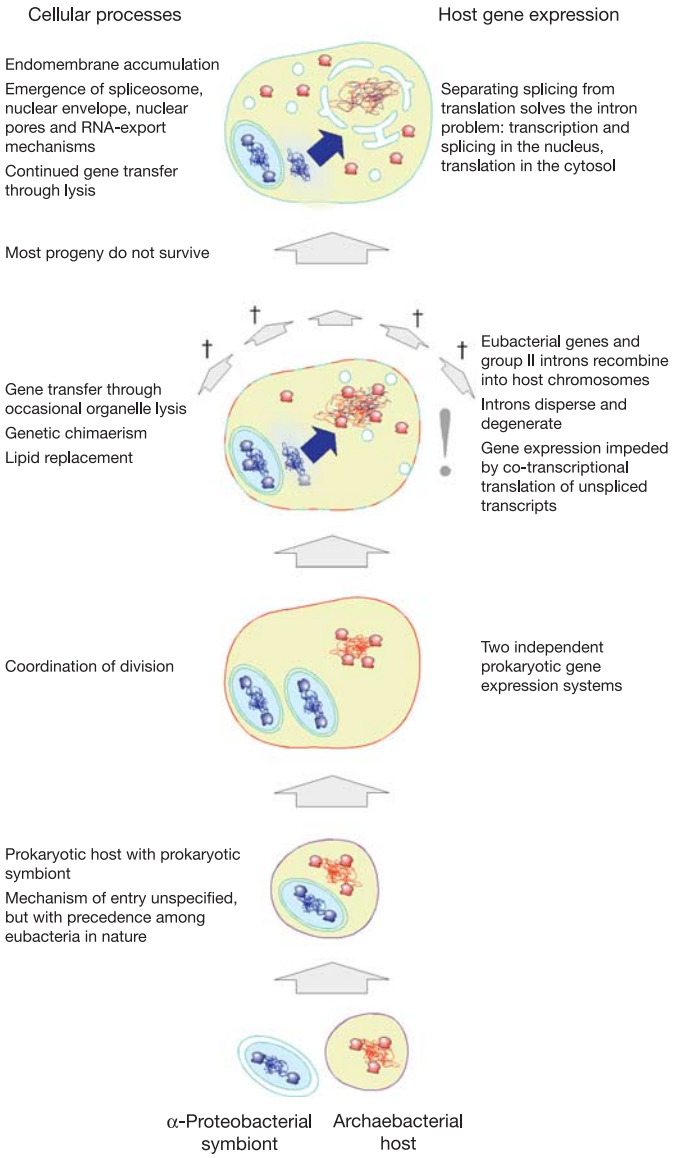

Figure 1 | Origin of nucleus–cytosol compartmentalization in the wake of mitochondrial origin. Blue arrows indicate symbiont-to-host gene transfer. The arrows marked with crosses symbolize the ill fate of most progeny that suffered intron invasion and other endosymbiont-triggered disturbances, resulting in a population bottleneck among progeny from a singular endosymbiotic event. Archaebacterial and eubacterial features are indicated in red and blue, respectively.

Introns in cytosolic chromosomes

Our salient considerations start with a simple idea proposed by Cavalier-Smith[14]: "…the slow splicing of spliceosomal introns would have made it difficult for them to evolve in bacteria (or in mitochondria or chloroplasts) because, if splicing is slow, the coexistence of DNA and functional ribosomes in the same cell compartment would allow ribosomes to translate unspliced premessengers and make incorrect proteins with intron sequences or, if the introns had stop codons, truncated and chimaeric proteins", as similarly expressed by Doolittle[32]. We require two additional premises for our inference.

As the simplest null hypothesis for the host that acquired the mitochondrion, we assume that it was a prokaryote—not a eukaryote, as previous authors have assumed[14,30,32]—and, therefore, that it lacked anything similar to a nucleus. In support of that premise, there are prokaryotes (albeit eubacterial) known to harbour eubacterial endosymbionts[33]. Thus, in contrast to some current views[30], the host need neither have been phagotrophic nor nucleate. Biologists agree that the host was in some way related to archaebacteria, although the nature of that relationship is still debated[34]. Either the host had a nucleus, or it did not. Our premise is that it did not; therefore, it was a prokaryote related to archaebacteria (thus an archaebacterium), which is compatible with any model for the origin of mitochondria that entails a prokaryotic host[35,36]. However, our present proposal is independent of the microbial physiology underlying that symbiosis. We also assume that mitochondria stem from α-proteobacteria[20,23,34–36]. Thus, without specifying a mechanism by which the mitochondrial endosymbiont came to reside within the cytosol of its prokaryotic host, but noting that such entry is possible[33], our inference starts with an α-proteobacterial ancestor of mitochondria and hydrogenosomes[20,22,23,35] inside an archaebacterial host[34–36] with one (or more) cytosolic chromosome(s) (Fig. 1, bottom).

This bipartite cell would not be an immediate success story: it would have nothing but problems instead. Rates of cell division for the host and the symbiont would have to reach comparable levels[37]. Only progeny that synchronized these rates would persist as consortia. If a host occasionally lyses, symbionts are set free. However, if a symbiont occasionally lyses, a genome's worth of eubacterial DNA is left in the host's cytosol, free to recombine. This situation is similar to Doolittle's ratchet[38] but differs in three salient points: the host here is a prokaryote, not a eukaryote; transfer occurs from a persistent resident symbiont, not from undigested meals; and gene transfer from that symbiont is facilitated by the lack of a nuclear membrane. As long as there is more than one symbiont per progeny, symbiont lysis can occur repeatedly, resulting in a constant flow of symbiont DNA into the chromosome(s) of the host.

Intron invasion causes problems

For the host, spreading group II introns that have hitchhiked into its chromosomes become an issue. In modern α-proteobacteria, group II introns occur at up to ~30 copies per genome[18]. Prokaryotic group II introns require the reverse transcriptase and maturase activities of their intron-encoded protein for mobility and splicing[17,18], have as yet unknown molecular control mechanisms, and are probably kept at bay by purifying selection in large prokaryotic populations[28,39]. They are usually—but not always—inserted either within intergenic spacers or within mobile elements, like insertion sequences, where they are not really introns but just other mobile elements[18,28]. Given the antiquity of many eukaryotic intron positions and the observed endonuclease-dependent mobility of modern group II introns in prokaryotes[17,18], the initial invasion of eukaryotic introns is most simply envisaged in the form of mobile, *bona fide* group II elements from the mitochondrion—their transition to spliceosome-dependent introns having occurred subsequently at the newly occupied sites. Spreading group II introns constitute a genetic burden, but not

an insurmountable one, because prokaryotes and organelles can effectively express genes that contain group II introns[18,28].

A problem of a much more severe nature arises, however, with the mutational decay of group II introns, resulting in inactivation of the maturase and/or RNA structural elements in at least some of the disseminated copies. Modern examples from prokaryotes and organelles suggest that splicing with the help of maturase and RNA structural elements provided by intact group II introns in *trans*[18] could have initially rescued gene expression at such loci, although maturase action in *trans* is much less effective than in *cis*[18]. Thus, the decay of the maturase gene in disseminated introns poses a requirement for invention of a new splicing machinery. However, as discussed below, the transition to spliceosome-dependent splicing will also impose an unforgiving demand for inventions in addition to the spliceosome.

Spliceosomes contain five snRNAs and about 200 proteins[27], and exist in two distinct forms for removing U2- and U12-dependent introns[28]. Evidence for an evolutionary transition of group II elements into spliceosomal introns comes from their similar splicing mechanisms[12,18,28]. Furthermore, the evolutionary conservation of the spliceosome[27] and intron positions[25,26] suggests that this transition occurred in the eukaryote ancestor, entailing the recruitment of group II-derived RNAs in *trans*, the precursors of snRNAs[18,28], plus novel accessory proteins[27].

The Sm-domain, a protein structural module involved in the still poorly characterized RNA-processing reactions in archaebacteria[40], was probably pivotal in that transition. The core complex of today's spliceosomes contains ~20 paralogous Sm-domain proteins that are conserved across eukaryotes[40]. It seems that the protospliceosome recruited the Sm-domain, possibly to replace the maturase, while retaining group II RNA domains (snRNAs) ancestrally germane to the splicing mechanism[18,28]. While the later evolution of the spliceosome entailed diversification with the recruitment of additional proteins[27]—leading to greater efficiency—the simpler, ancestral protospliceosome could, in principle, rescue expression of genes containing degenerate group II introns in a maturase-independent manner, but at the dear cost of speed.

Translation in prokaryotes is fast, of the order of 10 amino acids per second[41], whereas splicing by spliceosomes is slow, in the range of 0.005–0.01 intron per second in globins[42]. Although translation in modern eukaryotes is slower, ~1 amino acid per second[43], it is still much faster than modern splicing; because our assumed host is a prokaryote, its rate is relevant. The initial, as yet unoptimized protospliceosome—containing the precursors of *bona fide* snRNAs and still recruiting Sm-domain proteins[27,40]—was surely even slower than the modern one, bringing proper co-transcriptional translation of spliceosome-dependent genes virtually to a halt. Furthermore, in intron-bearing transcripts, the protospliceosome would have had trouble merely gaining access to the ribosome-covered mRNA: a problem that, in modern group II introns, is overcome by the maturase sequestering the ribosome-binding site[18]. Thus, ribosomes translating nascent transcripts that bear spliceosomal introns is an extremely unhealthy situation because few functional proteins will ensue[14,32], and the prospects of any descendants emerging from this situation are bleak. The only recognizable mechanism operating in favour of this clumsy chimaera is weakened purifying selection operating on its exceptionally small initial population[28,39].

Up to this point, the processes that generated introns and the spliceosome require no special inventions or unusual selective pressures; they stem solely from the polarity of the symbiosis, which governs the direction of gene transfer (Fig. 1). Finding a solution to the new problem of slow spliceosomes in the presence of fast and abundant ribosomes required an evolutionary novelty.

Solving the intron problem

There are three obvious routes for solving the problem of possessing spliceosome-dependent introns in co-transcriptionally translated mRNA. The first solution would be the spontaneous invention of an extremely fast and efficient spliceosome capable of outrunning the ribosomes. This entails a splicing efficiency in the ancestral spliceosome exceeding that of the modern one, and requires the preadaptive evolution of an unselected catalytic function (efficient splicing) in the first protospliceosome before the origin of its mRNA intron substrates, and therefore can be dismissed for that reason. The second is the invention of a mechanism to rapidly and efficiently remove spliceosomal introns from DNA, which—given eukaryote intron antiquity—apparently did not occur. The third solution is the invention of a means to physically separate splicing from translation, allowing the former (slow) process to occur to completion first, before the latter (fast) process sets in. Physical separation in cells usually entails membranes, so the third solution would involve the invention of a membrane separating splicing from translation, with pores sufficiently large and selective enough to export matured ribosomal subunits, mRNA and tRNA. But whence does the nuclear envelope originate?

In the modern cell cycle, the nuclear envelope is continuous with the endoplasmic reticulum[44] (ER), and their evolutionary origins are probably related[29,30]. For our present purpose, the existence of a primitive endomembrane consisting of a single eubacterial lipid bilayer is sufficient. The endomembrane system can assume any form, but only sheaths proximal to the host nucleoid will tend to spatiotemporally separate nascent transcripts from ribosomes. Any heritable variation promoting the exclusion of ribosomes from chromosomes until splicing is completed, releasing only processed transcripts to ribosomes in the cytosol, would enhance survival. Progeny that failed to physically separate mRNA processing from translation would not survive, nor would those that failed to invent pore complexes to allow chromosome–cytosol interaction[29,30].

The result of our inference is a cell bearing a mitochondrion, chimaeric chromosomes contained within a membrane-bounded splicing compartment (the nucleus), and archaebacterial ribosomes in the cytosol that translate mature mRNAs from intron-containing and intronless genes of both host and symbiont origin (Fig. 1, top). The invention of the nucleus was mandatory to allow the expression of intron-containing genes in a cell whose ribosomes were faster than its spliceosomes. With the original intron-encoded protein[18] no longer essential, its genome-wide mutational demise would have ended the initial intron invasion and left the ancestral intron set in place, the later emergence of introns in new positions notwithstanding[25,26].

A dedicated translation compartment

This view implicates a chromosome-free cytosol, not a nucleus, as the genuinely novel eukaryotic cell compartment: a dedicated translation compartment that is free of transcriptionally active chromosomes. The nucleoplasm is hardly novel in comparison to prokaryotes, because it has retained the job of chromosome maintenance and all basic gene expression functions other than translation. The evolutionary separation of translation from splicing required that mRNA, ribosomal subunits and tRNA mature in the nucleus, culminated by their export—as it occurs in modern eukaryotes, where mRNA export is coupled to splicing, with specific spliceosome components serving as mRNA export factors[45,46]. This is consistent with our proposal, as is the evidence that the eukaryotic Ntf2 (nuclear transport factor 2; also known as Nutf2) family, to which key proteins of mRNA export—Mex67/TAP/Nxf1 and Mtr2/Nxt1—belong, has homologues in α-proteobacteria[29].

In addition to splicing, eukaryotes possess elaborate mRNA surveillance mechanisms, in particular nonsense-mediated decay (NMD), to assure that only correctly processed mature mRNAs are translated, while aberrant mRNAs and those with premature termination codons are degraded[28,47,48]. The initial intron invasion would have precipitated a requirement for mechanisms to identify

exon junctions and to discriminate exons (with frame) from introns (without frame), as well as properly from improperly spliced transcripts. Thus, NMD might be a direct evolutionary consequence of newly arisen genes-in-pieces. Consistent with that view, the eukaryotic NMD machinery is composed of components derived from both the archaebacterial translation system and the eubacterial post-segregational cell-killing systems[49]. Thus, even if eukaryotes are found that have no introns left at all, they should possess mRNA surveillance mechanisms[47,48] as a remnant of their intron-laden past.

Reports that translation occurs to a substantial extent in the mammalian nucleus[50], the evidence for which remains controversial[47,48,51,52], are difficult to reconcile with our proposal. While many nuclear factors are involved in NMD[47,48], the participation, if any, of active nuclear protein synthesis in NMD remains uncertain[47,48]. The need to separate translation from mRNA maturation by the nuclear membrane would seem compelling for all intron-rich eukaryotes, both modern and ancestral.

Conclusion

Our suggestion for the origin of the nucleus differs from previous views on the topic, which either posit that the nuclear membrane was beneficial to (not mandatory for) its inventor by protecting chromosomes from shearing at division[30], or offer no plausible selective mechanism at all. The initial eukaryote we infer requires a nucleus only during phases of the cell cycle where genes are expressed, and thus is compatible with either an ancestral or a derived state for closed mitosis[53]. The archaebacterial nature of eukaryotic informational genes[34] involved in nucleocytoplasmic information storage and expression is readily accommodated by our proposal, as is the α-proteobacterial link with eukaryotic operational (biosynthetic) genes[34] of various cytosolic pathways[54]. While contributions from additional symbionts or other gene donors at the origin of the nucleus need not be excluded, our inference requires no more than two partners, each possessing a naturally diverse collection of genes, in order to account for available observations. Neither primitively amitochondriate eukaryotes nor eukaryotic-type spliceosomes in prokaryotes could be accommodated by the premises stated here. As with the spliceosome, we suggest that genes[55] and protein folds[56] specific to the eukaryotic lineage arose after the acquisition of mitochondria and before the divergence of the major eukaryotic lineages[57]. An important role for spreading introns in that phase of gene invention, involving rampant exon shuffling[1] mediated by intron sequences that were ancestrally homologous, is implicated.

1. Gilbert, W. Why genes in pieces? *Nature* **271,** 501 (1978).
2. Doolittle, W. F. Genes in pieces: Were they ever together? *Nature* **272,** 581–582 (1978).
3. Darnell, J. E. Implications of RNA—RNA splicing in evolution of eukaryotic cells. *Science* **202,** 1257–1260 (1978).
4. Doolittle, W. F. Revolutionary concepts in evolutionary cell biology. *Trends Biochem. Sci.* **5,** 147–149 (1980).
5. Darnell, J. E. & Doolittle, W. F. Speculations on the early course of evolution. *Proc. Natl Acad. Sci. USA* **83,** 1271–1275 (1986).
6. Cech, T. R., Zaug, A. J. & Grabowski, P. J. *In vitro* splicing of the ribosomal-RNA precursor of *Tetrahymena*—involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27,** 487–496 (1981).
7. Gilbert, W. The RNA world. *Nature* **319,** 618 (1986).
8. Gilbert, W. The exon theory of genes. *Cold Spring Harb. Symp. Quant. Biol.* **52,** 901–905 (1987).
9. Stoltzfus, A., Spencer, D. F., Zuker, M., Logsdon, J. M. & Doolittle, W. F. Testing the exon theory of genes: The evidence from protein structure. *Science* **265,** 202–207 (1994).
10. Poole, A., Jeffares, D. & Penny, D. Prokaryotes, the new kids on the block. *Bioessays* **21,** 880–889 (1999).
11. Forterre, P. & Philippe, H. Where is the root of the universal tree of life? *Bioessays* **21,** 871–879 (1999).
12. Jeffares, D. C., Poole, A. M. & Penny, D. Relics from the RNA world. *J. Mol. Evol.* **46,** 18–36 (1998).
13. Cech, T. R. The generality of self-splicing RNA: Relationship to nuclear mRNA splicing. *Cell* **44,** 207–210 (1986).
14. Cavalier-Smith, T. Intron phylogeny: A new hypothesis. *Trends Genet.* **7,** 145–148 (1991).
15. Ferat, J.-L. & Michel, F. Group II self splicing introns in bacteria. *Nature* **364,** 358–361 (1993).
16. Roger, A. J. & Doolittle, W. F. Why introns-in-pieces? *Nature* **364,** 289–290 (1993).
17. Cousineau, B., Lawrence, S., Smith, D. & Belfort, M. Retrotransposition of a bacterial group II intron. *Nature* **404,** 1018–1021 (2000); correction **414,** 84 (2001).
18. Lambowitz, A. M. & Zimmerly, S. Mobile group II introns. *Annu. Rev. Genet.* **38,** 1–35 (2004).
19. Nixon, J. E. *et al.* A spliceosomal intron in *Giardia lamblia*. *Proc. Natl Acad. Sci. USA* **99,** 3701–3705 (2002).
20. Tovar, J. *et al.* Mitochondrial remnant organelles of *Giardia* function in iron–sulphur protein maturation. *Nature* **426,** 172–176 (2003).
21. Simpson, A. G., MacQuarrie, E. K. & Roger, A. J. Early origin of canonical introns. *Nature* **419,** 270 (2002).
22. Vanacova, S., Yan, W., Carlton, J. M. & Johnson, P. J. Spliceosomal introns in the deep-branching eukaryote *Trichomonas vaginalis*. *Proc. Natl Acad. Sci. USA* **102,** 4430–4435 (2005).
23. Hrdy, I. *et al. Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. *Nature* **432,** 618–622 (2004).
24. van der Giezen, M. & Tovar, J. Degenerate mitochondria. *EMBO Rep.* **6,** 525–530 (2005).
25. Rogozin, I. B., Wolf, Y. I., Sorokin, A. V., Mirkin, B. G. & Koonin, E. V. Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr. Biol.* **13,** 1512–1517 (2003).
26. Roy, S. W. & Gilbert, W. Rates of intron loss and gain: Implications for early eukaryotic evolution. *Proc. Natl Acad. Sci. USA* **102,** 5773–5778 (2005).
27. Collins, L. & Penny, D. Complex spliceosomal organization ancestral to extant eukaryotes. *Mol. Biol. Evol.* **22,** 1053–1066 (2005).
28. Lynch, M. & Richardson, A. O. The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* **12,** 701–710 (2002).
29. Mans, B. J., Anantharaman, V., Aravind, L. & Koonin, E. V. Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* **3,** 1612–1637 (2004).
30. Cavalier-Smith, T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52,** 297–354 (2002).
31. Staub, E., Fiziev, P., Rosenthal, A. & Hinzmann, B. Insights into the evolution of the nucleolus by an analysis of its protein domain repertoire. *Bioessays* **26,** 567–581 (2004).
32. Doolittle, W. F. The origin of introns. *Curr. Biol.* **1,** 145–146 (1991).
33. von Dohlen, C. D., Kohler, S., Alsop, S. T. & McManus, W. R. Mealybug β-proteobacterial symbionts contain γ-proteobacterial symbionts. *Nature* **412,** 433–436 (2001).
34. Rivera, M. C. & Lake, J. A. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431,** 152–155 (2004).
35. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392,** 37–41 (1998).
36. Vellai, T., Takács, K. & Vida, G. A new aspect on the origin and evolution of eukaryotes. *J. Mol. Evol.* **46,** 499–507 (1998).
37. Finlay, B. J. & Fenchel, T. An anaerobic ciliate as a natural chemostat for the growth of endosymbiotic methanogens. *Eur. J. Protistol.* **28,** 127–137 (1992).
38. Doolittle, W. F. You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* **14,** 307–311 (1998).
39. Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302,** 1401–1404 (2003).
40. Khusial, P., Plaag, R. & Zieve, G. W. LSm proteins form heptameric rings that bind to RNA via repeating motifs. *Trends Biochem. Sci.* **30,** 522–528 (2005).
41. Sorensen, M. A., Kurland, C. G. & Pedersen, S. Codon usage determines translation rate in *E. coli*. *J. Mol. Biol.* **207,** 365–377 (1989).
42. Audibert, A., Weil, D. & Dautry, F. *In vivo* kinetics of mRNA splicing and transport in mammalian cells. *Mol. Cell. Biol.* **22,** 6706–6718 (2002).
43. Palmiter, R. D. Quantitation of parameters that determine the rate of ovalbumin synthesis. *Cell* **4,** 189–197 (1975).
44. Levine, T. & Rabouille, C. Endoplasmic reticulum: One continuous network compartmentalized by extrinsic cues. *Curr. Opin. Cell Biol.* **17,** 362–368 (2005).
45. Reed, R. & Hurt, E. A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108,** 523–531 (2002).
46. Luo, M. L. *et al.* Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* **413,** 644–647 (2001).
47. Fasken, M. B. & Corbett, A. H. Process or perish: Quality control in mRNA biogenesis. *Nature Struct. Mol. Biol.* **6,** 482–488 (2005).
48. Maquat, L. E. Nonsense mediated mRNA decay: Splicing, translation and mRNP dynamics. *Nature Rev. Mol. Cell Biol.* **5,** 89–99 (2004).
49. Anantharaman, V. & Aravind, L. New connections in the prokaryotic toxin–antitoxin network: Relationship with the eukaryotic nonsense-mediated RNA decay system. *Genome Biol.* **4,** R81 (2003).

50. Iborra, F. J., Jackson, D. A. & Cook, P. R. Coupled transcription and translation within nuclei of mammalian cells. *Science* **293,** 1139–1142 (2001).
51. Dahlberg, J. E. & Lund, E. Does protein synthesis occur in the nucleus? *Curr. Opin. Cell Biol.* **16,** 335–338 (2004).
52. Cosson, B. & Philippe, M. Looking for nuclear translation using *Xenopus* oocytes. *Biol. Cell.* **95,** 321–325 (2003).
53. Heath, I. B. Variant mitoses in lower eukaryotes: Indicators of the evolution of mitosis? *Int. Rev. Cytol.* **64,** 1–80 (1980).
54. Gabaldón, T. & Huynen, M. A. Reconstruction of the proto-mitochondrial metabolism. *Science* **301,** 609 (2003).
55. Makarova, K. S., Wolf, Y. I., Mekhedov, S. L., Mirkin, B. G. & Koonin, E. V. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* **33,** 4626–4638 (2005).
56. Yang, S., Doolittle, R. F. & Bourne, P. E. Phylogeny determined by protein domain content. *Proc. Natl Acad. Sci. USA* **102,** 373–378 (2005).
57. Adl, S. M. *et al.* The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* **52,** 399–451 (2005).

Acknowledgements We thank the DFG (W.M.) and the NIH intramural research program (E.V.K.) for financial support, and M. Embley, T. Dagan, I. Rogozin, T. Senkevich and Y. Wolf for discussions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to W.M. (w.martin@uni-duesseldorf.de).