observation led to the study of the (L)-α-threofuranosyl-(2′→3′) oligonucleotide system (Fig. 1, bottom right), which, according to recent observations, is in fact a weak pairing system [see (*47*)].
40. I. Schlönvogt *et al.*, *Helv. Chim. Acta* **79**, 2316 (1996).
41. R. Micura, R. Kudick, S. Pitsch, A. Eschenmoser, *Angew. Chem. Int. Ed. Engl.* **38**, 680 (1999); R. Micura, M. Bolli, N. Windhab, A. Eschenmoser, *ibid.* **36**, 870 (1997).
42. M. Bolli, R. Micura, S. Pitsch, A. Eschenmoser, *Helv. Chim. Acta* **80**, 1901 (1997).
43. M. Renz, R. Lohrmann, L. E. Orgel, *Biochim. Biophys. Acta* **240**, 463 (1971); see also (*12*).
44. R. Micura, M. Bolli, A. Eschenmoser, unpublished work.
45. M. Bolli, R. Micura, A. Eschenmoser, *Chem. Biol.* **4**, 309 (1997).
46. The data (except those for the hexose series) were determined for duplexes containing thymine instead of uracil as a nucleobase (due to the protection group strategy used in the synthesis of the pentopyranosyl members).
47. K.-U. Schöning, P. Scholz, A. Eschenmoser, unpublished work.
48. T. Vivlemore, S. Pitsch, A. Eschenmoser, unpublished work.
49. The T_m-value for the $r(A_8) \cdot r(T_8)$-duplex given in table 1 and figure 3 of (*36*) mistakenly refers to a concentration of 3 μM instead of the standard 10 μM.
50. L. E. Orgel, *Nature* **358**, 203 (1992).
51. ——— and R. Lohrmann, *Acc. Chem. Res.* **7**, 368 (1974); L. E. Orgel, *ibid.* **28**, 109 (1995); T. Wu and L. E. Orgel, *J. Am. Chem. Soc.* **114**, 317(1992); *ibid.*, p. 5496; *ibid.*, p. 7963.
52. M. Kurz, K. Göbel, C. Hartel, M. W. Göbel, *Angew. Chem. Int. Ed. Engl.* **36**, 842 (1997).
53. G. Ertem and J. P. Ferris, *Nature* **379**, 238 (1996).
54. T. Li and K. C. Nicolaou, *ibid.* **369**, 218 (1994); K. D. James and A. D. Ellington, *Chem. Biol.* **4**, 595 (1997).
55. G. v. Kiedrowski, *Angew. Chem. Int. Ed. Engl.* **25**, 932 (1986); D. Sievers and G. v. Kiedrowski, *Chem. Eur. J.* **4**, 629 (1998); A. Luther, R. Brandsch, G. v. Kiedrowski, *Nature* **396**, 245 (1998).
56. Of great interest in the etiological context, however, are recent demonstrations that, in contrast to previous belief, auto- and cross-catalytic molecular replication with high turnover numbers can occur in the oligopeptide series [D. H. Lee, J. R. Granja, J. A. Martinez, K. Severin, M. R. Ghadiri, *Nature* **382**, 525 (1996); D. H. Lee, K. Severin, Y. Yokobayashi, M. R. Ghadiri, *ibid.* **390**, 591 (1997); see also S. Yao, I. Ghosh, R. Zutshi, J. Chmielewski, *ibid.* **396**, 447 (1998)]. For work on autocatalytic replication in other chemical systems, see T. Tjirikua, P. Ballester, J. Rebek Jr., *J. Am. Chem. Soc.* **112**, 1249 (1990); E. A. Wintner, M. M. Conn, J. Rebek Jr., *ibid.* **116**, 8877 (1994); K. Soai, T. Shibata, H. Morioka, K. Choji, *Nature* **378**, 767 (1995).
57. J. A. Doudna and J. W. Szostak, *Nature* **339**, 519 (1989); A. A. Beaudry and G. F. Joyce, *Science* **257**, 635 (1992); E. H. Ekland and D. P. Bartel, *Nature* **382**, 373 (1996).
58. J. Oró, *Nature* **191**, 1193 (1961); J. P. Ferris and L. E. Orgel, *J. Am. Chem. Soc.* **88**, 1074 (1966); J. P. Ferris and W. J. Hagan Jr., *Tetrahedron* **40**, 1093 (1984) (review).
59. G. F. Joyce, *Nature* **338**, 217 (1989).
60. On the other hand, substantial progress has recently been made with regard to the formation of (3′→5′)/(2′→5′)-ribofuranosyl oligonucleotide mixtures by mineral-assisted oligocondensation of activated 5′ mononucleotides [J. P. Ferris and G. Ertem, *J. Am. Chem. Soc.* **115**, 12270 (1993); J. P. Ferris, A. R. Hill Jr., R. Liu, L. E. Orgel, *Nature* **381**, 59 (1996); (*53*)].
61. Of possible relevance in this context are the preference for ribose formation in kinetically controlled aldolization reactions [D. Müller *et al.*, *Helv. Chim. Acta* **73**, 1410 (1990); G. Zubay, *Origins Life Evol. Biosphere* **28**, 13 (1998)] and the fixation of the furanose form of ribose by preferential functionalization of the sterically least hindered primary hydroxyl group [B. Bennua-Skalmowski, K. Krolikiewicz, H. Vorbrüggen, *Tetrahedron Lett.* **36**, 7845 (1995); J. Wengel and C. Scheuer-Larsen, *Bioorg. Med. Chem. Lett.* **7**, 1999 (1997)] or by annelation of the ribofuranose ring by cis-fused five-membered ring α-diol derivatives, for example, cyclophosphates (C. Spinner, S. Pitsch, R. Krishnamurthy, A. Eschenmoser, unpublished work).
62. Although the numerous nucleic acid analogs synthesized recently in antisense oligonucleotide chemistry [for reviews, see J. Hunziker and C. Leumann, in *Modern Synthetic Methods*, B. Ernst and C. Leumann, Eds. (Verlag Helvetica Chimica Acta, Basel, Switzerland, 1995), vol. 7, p. 331; B. Hyrup and P. E. Nielsen, *Bioorg. Med. Chem. Lett.* **4**, 5 (1996); (*29*)] are an additional illustration of the remarkably broad spread of the nucleobase-pairing potential among organic oligomer structures, the backbones of most of these systems do not correspond to the criteria adopted here for potentially natural nucleic acid alternatives. In a broader etiological context, however, pairing systems such as Nielsen's "polyamide nucleic acid" [P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **254**, 1497 (1991); M. Koppitz, P. E. Nielsen, L. E. Orgel, *J. Am. Chem. Soc.* **120**, 4563 (1998)] or amino acid–based systems of the type described by, for example, U. Diederichsen [*Angew. Chem. Int. Ed. Engl.* **35**, 445 (1996)] or M. Fujii *et al.* [*Chem. Commun.* **1998**, 717 (1998)] may have to become part of the range of structure types that must be screened with respect to their accessibility under natural conditions.
63. The author is deeply indebted to R. Lerner (TSRI, La Jolla) and to R. Hütter (ETH, Zürich) for their support. I thank my present or former collaborators O. Jungmann, H. Wippo, M. Stanek, H. Huynh, R. Kudick, F. Reck, G. Ceulemans, and M. Bolli at TSRI; K.-U. Schöning, P. Scholz, T. Vivlemore, S. Pitsch, and R. Micura at ETH for providing unpublished results mentioned in this article; and R. M. Wolf (Novartis Pharma AG, Basel, Switzerland) for producing Fig. 4. The other figures were drawn by R. Krishnamurthy, to whom I express my appreciation for his assistance in operating the La Jolla branch of my research group. I thank S. Weinberg (University of Texas, Austin) for the source of the Einstein quote; I also thank colleagues who helped with the translation. Finally, I am grateful to C. Wintner (Haverford College, Haverford, PA), who critically read the manuscript, made valuable suggestions, and improved the English.

REVIEW

Phylogenetic Classification and the Universal Tree

W. Ford Doolittle

From comparative analyses of the nucleotide sequences of genes encoding ribosomal RNAs and several proteins, molecular phylogeneticists have constructed a "universal tree of life," taking it as the basis for a "natural" hierarchical classification of all living things. Although confidence in some of the tree's early branches has recently been shaken, new approaches could still resolve many methodological uncertainties. More challenging is evidence that most archaeal and bacterial genomes (and the inferred ancestral eukaryotic nuclear genome) contain genes from multiple sources. If "chimerism" or "lateral gene transfer" cannot be dismissed as trivial in extent or limited to special categories of genes, then no hierarchical universal classification can be taken as natural. Molecular phylogeneticists will have failed to find the "true tree," not because their methods are inadequate or because they have chosen the wrong genes, but because the history of life cannot properly be represented as a tree. However, taxonomies based on molecular sequences will remain indispensable, and understanding of the evolutionary process will ultimately be enriched, not impoverished.

The impulse to classify organisms is ancient, as is the desire to have classification reflect the "natural order." Before Darwin, biologists thought that God or some other eternal principle created that order (*1*, *2*). After Darwin (*3*), they knew the ordering principle to be shared descent from an ever more limited number of common ancestors (Fig. 1), back to the last common ancestor of all living things. A phylogenetic classification is thus the only natural one, and it should be "inclusively hierarchical" (*1*): Each species should be part of one and only one genus, each genus should be part of one and only one family, and so forth.

Much of modern phylogenetics is molecular phylogenetics. Microbial phylogeneticists in particular depend on molecular sequence characters, because prokaryotes (Bacteria and Archaea) offer relatively little in the way of complex morphology and behavior. Beyond this practical consideration is the understanding that molecular sequences define, in the words of Zuckerkandl and Pauling, "the essence of the organism"—not only do genes reveal the phy-

Canadian Institute for Advanced Research, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada. E-mail: ford@is.dal.ca

logenetic pattern, they engender and embody it (*4*).

Since 1965, when arguments in favor of molecular phylogenetics were first advanced, gene sequence data have become astonishingly abundant. Today, molecular phylogeneticists appear to have realized Darwin's hope for a universal phylogenetic tree (*5*), a hierarchical classification of "groups subordinate to groups" going back to the first dawn of life, when all life was microbial. This tree is shown in cartoon form, emphasizing only its early branchings, in Fig. 2.

Establishment of a Universal Molecular Phylogeny

In its early period (1965–77), microbial molecular phylogeny depended on sequences of proteins, in particular, ferredoxins and cytochromes. These data delineated certain relationships among bacteria and gave strong support to the "serial endosymbiosis theory"—the notion that mitochondria and chloroplasts are descendants of what are now called α-proteobacteria and cyanobacteria, respectively (*6*).

In the mid-1970s, Woese and his collaborators began to assemble the massive database of sequence information on small subunit ribosomal RNA (SSU rRNA) on which the current universal tree rests (*5, 7*). This molecule is superior to cytochromes or ferredoxins as a "molecular chronometer" for many reasons, spelled out by Woese at the outset. It is abundant, it is coded for by organellar as well as nuclear and prokaryotic genomes, it has slow- and fast-evolving portions ("hour hands and minute hands"), and it has a universally conserved structure (*7*). Two other factors that contribute materially to the confident use of SSU rRNA are its obviously ancient and essential fundamental function in the cellular economy and its interaction with many (well over 100) other coevolved cellular RNAs and proteins (*8*). These last features would seem to make rRNA genes the least liable of all genes to experience interspecific lateral gene transfer (LGT).

Figure 2 presents a crude sketch of the universal SSU rRNA tree (*5*), commonly taken as a representation of organismal phylogeny and the basis for a natural classification. The distinct and cohesive nature of each of its three "domains" (Archaea, Bacteria, and Eukarya) and the branching pattern of hundreds of subordinate taxa (kingdoms and lower divisions) within each domain are supported by SSU rRNA sequences. The primary branching pattern (separating Bacteria on one side from Archaea and Eukarya on the other) is not. For nontrivial technical reasons (*9*), this "rooting" rests on analyses of a few families of ubiquitous duplicated protein-coding genes.

How True Is This Phylogeny?

There is much support for the general features of Fig. 2. Many other molecular phylogenies and some strong phenotypic characters distinguish its three domains and concur in supporting major divisions within them (*10*). For instance, all archaea use diphytanylglycerol diether or dibiphytanyldiglycerol tetraether or both as major lipid constituents, whereas bacteria and eukaryotes use diacylglycerol-derived lipids; bacteria with cell walls employ petidoglycan as a strengthening agent, but archaea and eukaryotes never do; eukaryotes all have tubulin- and actin-based cytoskeletons, whereas bacteria and archaea have only very distant homologs of these proteins and no cytoskeletons in the eukaryotic sense. The specific affinity (sisterhood) between archaea and eukaryotes shown in Fig. 2 is also supported by the strikingly eukaryotic nature of the components and mechanisms of archaeal replication, transcription, and translation systems (*11*).

At the same time, there is now less general agreement about the larger meaning and truth of Fig. 2 than there would have been even a year ago. This contradictory state of affairs has two causes. First, more critical analyses of both rRNA and protein-based phylogenies show that artifacts related to within-molecule and between-lineage differences in evolutionary rate and mutational saturation can be misleading about deep branchings and the rooting of the tree (*12–14*). Second, and completely independent of these methodological problems, are doubts stemming from the fact that many genes give believably different phylogenies for the same organisms (*15, 16*), almost certainly because they have been "laterally transferred." If instances of LGT can no longer be dismissed as "exceptions that prove the rule," it must be admitted (i) that it is not logical to equate gene phylogeny and organismal phylogeny and (ii) that, unless organisms are construed as either less or more than the sum of their genes, there is no unique organismal phylogeny. Thus, there is a problem with the very conceptual basis of phylogenetic classification.

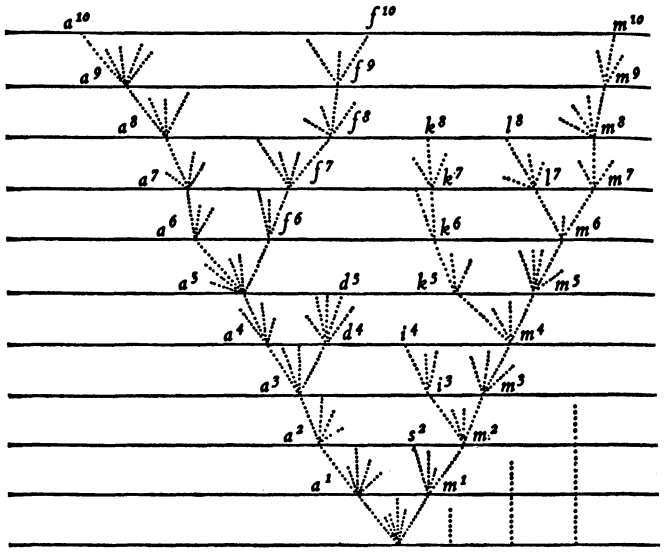

Fig. 1. Part of the only figure in the *Origin of Species*. Darwin first uses it to represent the divergence of variants within a species, showing successively more difference in a single lineage (a^1 through a^{10}) and splitting into multiple lineages (*m*, *s*, *i*, and so forth), some of which will become new species. Later, he expands the tree metaphor, explaining that "limbs divided into great branches ... were themselves once, when the tree was small, budding twigs; and this connection of the former and present buds by ramifying branches may well represent the classification of all extinct and living species in groups subordinate to groups" (*3*, p. 171).

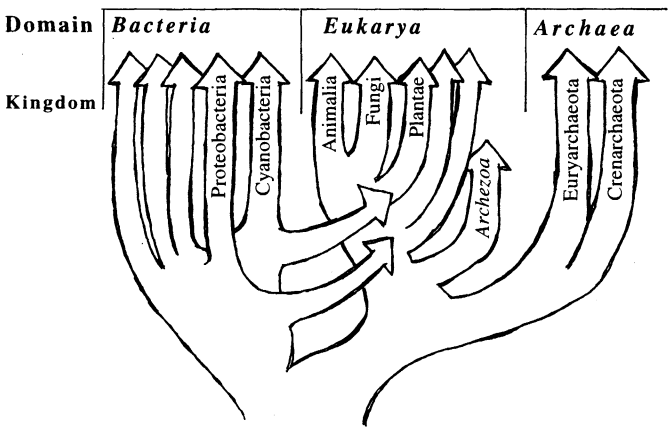

Fig. 2. The current consensus or standard model. Only a few of the "kingdoms" of the "domain" Bacteria are shown. Branching orders of several kingdoms within Bacteria and Eukarya remain in dispute. Mitochondrial and chloroplast endosymbioses are indicated by lower and upper diagonal arrows, respectively. Archezoa, as a subkingdom composed of primitively amitochondriate protists, may be extinct. For SSU rRNA trees with much more detail, see (*5*).

Methodological Problems

Concerns about resolution in deep phylogeny have come to the fore because of convincing demonstrations that SSU rRNA has, in a few striking instances, been unreliable. In eukaryote phylogeny, microsporidia are the exemplar. These anaerobic protists had been classified by Cavalier-Smith as "archezoa"—eukaryotes that diverged from the main line before the acquisition of mitochondria—on the basis of what looked to be primitive cytological features (*17*). Although a deep branching of microsporidia is favored by SSU rRNA analyses (Fig. 2) (*18*), protein phylogenies, most notably for the largest subunit of RNA polymerase, position microsporidia with or within the fungi (*14*). RNA polymerase is as fundamental to cell function and as integrated in its action with other components as is rRNA. There is no compelling reason other than pride of place to chose rRNA as the more reliable molecular chronometer.

Such inconsistencies have prompted a reexamination of SSU rRNA phylogenies and protein-sequence data (principally for elongation factors) that support all deep eukaryotic and prokaryotic branchings (*12–14, 19*). Philippe and colleagues (*12*) are the most enthusiastically deconstructive of the "revisionist" phylogeneticists. Not only do they assert that available methods are inadequate for reconstructing early evolution, but also that the rooting of the universal tree is hopelessly compromised by methodological artifacts and LGT.

Despite the vigor of their critique, even Philippe and colleagues (*12*) have not given up on deep phylogeny. They propose a fuller development of Fitch's concomitantly varying codon (covarion) theory [which predicts lineage-specific patterns of rate variation among sites (*20*)], more suitable taxon selection (especially of out-groups), and the identification of rare molecular events that unite taxa (such as a 12–amino acid insertion with semiconserved sequence found only in the elongation factor EF-1αs of animals, fungi, and microsporidia). For early eukaryote evolution, paralogous genes produced by duplications occurring since the origin of eukaryotes might also be fruitfully used. Deeper (preduplication) taxa would have single copies of such genes, identifiable as out-groups to the paralogous duplicates. A third approach would be to use events of LGT as characters in phylogenetic reconstruction, as Gogarten (*15*) has proposed: Taxa sharing a transferred gene must surely share a common (pretransfer) ancestor.

LGT Challenges the Conceptual Basis of Phylogenetic Classification

The endosymbiont hypothesis, widely accepted since the mid-1970s, imagined that although most of the α-proteobacterial endosymbiont's genes have been lost, some were transferred to their host's nuclear genome (*21*). Gene transfer from mitochondrion to nucleus is, of course, a form of LGT. Most biologists have nevertheless been comfortable seeing eukaryotes as a natural group, closest to archaea, probably because (i) the mitochondrion could be viewed as an invader, (ii) the genes transferred from it to the nuclear genome were thought to be relatively few in number, and (iii) the products of these genes were thought mostly to be targeted (by suitable leader sequences) back into mitochondria, serving their original functions.

However, modern surveys of eukaryotic gene phylogenies (*22*) show that many (perhaps most) enzymes involved in eukaryotic cytosolic metabolism are also of bacterial origin, not of archaeal ancestry as one would expect (Fig. 2). Resolution is, in most cases, inadequate to pinpoint bacterial sources. Most of these genes might have the same original protomitochondrial ancestry as those whose products still function in that organelle, and there is (as yet) no reason to suppose that multiple independent LGTs have played a major role in the evolution of eukaryotes since their origin. Nevertheless, serious modifications of the original endosymbiont hypothesis are called for (*23*), and one must ask why archaea are still considered to be the eukaryotes' closest relatives, when only a minority of eukaryotic genes may show this to be true.

Recent evidence that LGT is a major and continuing force in archaeal and bacterial evolution is dramatic and of three distinct sorts: analyses of guanine plus cytosine (GC) content and codon usage in individual genomes, genome-by-genome content comparisons, and individual gene trees. Lawrence and Ochman (*24*, p. 9413), from an analysis of GC content and codon usage in the completed *Escherichia coli* genome, concluded that an astonishing "755 of 4288 [open reading frames] ORFs [18%] have been introduced into the *E. coli* genome in at least 234 lateral transfer events since this species diverged from the *Salmonella* lineage 100 million years ago." Among these ORFs are determinants of all phenotypic characters (such as lactose utilization, citrate utilization, indole production, and propanediol utilization), which distinguish *E. coli* from *Salmonella enterica*.

Comparative examination of the gene contents of other completed genomes show that extensive LGT is not just an evolutionary peculiarity of enteric bacteria. For instance, the completed genome (*25*) of *Archaeoglobus fulgidis*, the heterotrophic archaeon often found metabolizing deep-sea oil supplies, bears many genes for fatty acid degradation that are unknown in other sequenced archaeal genomes, but recognizable as genes for fatty acid metabolism because they have homologs in bacteria. One might assert that these "bacterial" genes were present in the bacterial/archaeal common ancestor and lost in other archaea, but using this strategem in all such cases produces an ancestor whose enormous genome contained direct antecedents of all genes found in all contemporary prokaryotes. Moreover, there are many notable instances [3-hydroxy-3-methylglutaryl coenzyme A (HMGCoA) reductases of *Archaeoglobus*, lysyl–transfer RNA (tRNA) synthetase of spirochetes (*26*)] where interdomain transfer has clearly resulted in the displacement of an isofunctional resident enzyme.

Complete genome surveys have also allowed genome-versus-genome comparisons of genes present in both bacterial and archaeal domains. From analyses of four bacterial and two archaeal genomes, Lake and co-workers (*27*) have extracted 40 data sets for such genes: Whereas 6 of 12 replication, transcription, or translation-related genes support the topology of Fig. 2, only 1 of 28 trees for central metabolism and housekeeping functions shows this topology. Other surveys (*22*), with other methods, support the same general result: There has been extensive sharing (LGT) between bacterial and archaeal domains, especially of housekeeping biosynthetic or catabolic genes. Gene sharing within domains is less readily detectable in such crude surveys but surely is even more frequent.

LGT among and between bacteria and archaea is not new: Gogarten has been thoroughly documenting individual (especially prokaryotic) instances for several years, suggesting in 1993 (*28*) that, for many genes, "the tree of life becomes a net." Martin and collaborators have similarly been emphasizing the role of LGT in the evolution of eukaryotic metabolism (*16*). Much earlier, when the involvement of plasmids in the spread of antibiotic resistance among infectious bacteria was first understood, several authors expounded the evolutionary importance of LGT. Most outspoken were Sonea and Reanney (*29*), asserting that the activities of plasmids, phages, and other DNA exchange devices made all the planet's bacteria into a single "global superorganism."

These radical claims did not divert the mainstream of microbial evolutionary discourse. Plasmid-borne determinants seemed mostly restricted to genes for resistance to antibiotics and toxins or for use of unusual substrates—"dispensable" functions not at the core of any organism's biology. But, now LGT is known to have been the source of a substantial fraction of many bacterial and archaeal genomes, and it is known to have affected genes that are very much a part of the cellular economy, such as archaeal HMGCoA reductase, glutamine synthetase, Hsp^{70}, H^+-dependent adenosine triphosphatases (ATPases), and aminoacyl-tRNA synthetases (*26, 28*).

How Can a Phylogenetic Classification Be Preserved?

Can LGT still be treated as just a nuisance in phylogenetic classification, or is it the es-

sence of the phylogenetic process (at least for prokaryotes and the earliest eukaryotes) and thus a threat to the whole enterprise of classification? There are several popular and reasonable defenses for the conservative view that LGT is interesting but not a threat. I consider three, with arguments for and against.

Gene transfer seems unlikely to affect genes for replication, transcription, and translation, especially rRNA genes. LGT describes an outcome (incongruent gene phylogenies), not a specific biological process. In general, it might be thought that most processes producing incongruent phylogenies involve a "donor," which contributes only a small amount of DNA, and a recipient. It is presumably the lineage of the recipient cell (organism) and its progeny that should be represented by phylogenies such as Fig. 2. Those who argue for the untransferability of transcription- and translation-related genes are arguing that such genes reliably track that cellular (organismal) lineage.

There are two rationales for arguing in this way. First, transcription and translation genes are central to the "essence of the organism." They encode the hardware that reads the exchangeable genes for the cellular software (metabolism) and thus are more "fundamental" to the cell. Second, transcription and translation machinery are complex, and their individual components in any cellular lineage must have many highly coevolved interactions with each other: They should not integrate well with the components of substantially unrelated cellular lineages (*7*).

Against the first of these commonsense views, one might object that cells do not actually know what is fundamental to them, which of their genes encode hardware rather than software. Countering the second view, one could point out that some transcription/translation components (for instance, aminoacyl-tRNA synthetases) interact with few macromolecular partners, whereas complex multicomponent machines are also involved in many noninformational cellular processes (chaperonins, proteosomes, and adenosine triphosphatases). Also, there are examples of LGT affecting cellular hardware: (i) the replacement of a bacterial-type RNA polymerase by a phage-type enzyme in mitochondrial evolution (*30*), (ii) exchanges of mitochondrial elongation factor–G (EF-G) genes between eukaryotic nuclei and spirochetes (*31*), (iii) LGTs of ribosomal protein genes (*32*), and (iv) frequent between-domain exchanges of genes encoding aminoacyl-tRNA synthetases (*33*).

Nevertheless, the most trusted molecular chronometer, SSU rRNA, is at the very core of the cell's most complex machine. Surely, SSU rRNA genes are immune to LGT. Perhaps, but there are reasons, both old and new, to suspect that even rRNA genes can be transferred (*34*). Three decades ago, in vitro experiments showed that rRNA and proteins from very different bacteria can form partially functional ribosomes (*35*), and comparative structural studies continue to show that it is those nucleotide residues involved in intramolecular, not intermolecular, interactions that change most rapidly in evolution [providing the phylogenetic signal (*36*)]. Very recently, Squires and co-workers (*37*) have demonstrated that the SSU rRNA of *E. coli* can be completely replaced by that of *Proteus vulgaris* (and the ribosomal protein L11 binding domain of *E. coli* 23*S* can be replaced by the homologous region of yeast 28*S*) without reducing growth rate by more than 10 to 30%. Much smaller rates might easily be selected for against a no-growth background imposed by RNA-targeted antibiotics. Gupta (*38*) has not unreasonably suggested that such antibiotic pressure might drive LGT of SSU rRNA genes, as it has clearly done for aminoacyl-tRNA synthetase genes (*39*). Finally, Ueda *et al.* (*40*) have proposed that LGT may be a good explanation for the SSU rRNA heterogeneity observed within strains of *Streptomyces*.

For whatever reason, most genes will tell the same story as rRNA, even though LGT "noise" is higher than expected. The argument against this often-articulated view is that several preliminary genome-by-genome studies show it to be false (*22*). A thorough gene-by-gene analysis of all available genomes has yet to be presented, and some apparent cases of LGT will surely turn out to be methodological artifacts. It might still be the case that there will be more genes that support a division of living things into Bacteria, Archaea, and Eukarya than support any other single trifurcation (or other simple division) of all known taxa. Nevertheless, such a "majority rule" classification is not the "natural" scheme that Darwin, Zuckerkandl and Pauling, or Woese first had in mind. Inclusive organismal hierarchy may just not be a biological reality.

LGT was a problem in early evolution, but things have improved since. In a recent article (*41*) on "the universal ancestor," Woese reaffirmed and expanded his views of the last common ancestor, which he has held since his original discovery of the trifurcation in rRNA trees. He envisions a different tempo and mode of evolution, driven by LGT between primitive cells with as yet inefficient and error-prone replication, transcription, and translation and short exchangeable "operonal" chromosomes. Since then (after the divergence of Bacteria, Archaea, and Eukarya) genomes began to "anneal," becoming refractory to LGT as hardware components became refined and highly interdependent.

Inefficient and error-ridden primitive cells surely did once exist, but the patterns of prokaryotic gene trees (Fig. 3) can probably be accounted for by invoking LGT at the frequency inferred by Lawrence and Ochman (*24*) for *E. coli*'s past 100 million years, operating between cells not radically different from modern bacteria and archaea over the past 3.5 billion years, which is the age of the earliest cellular fossils. (LGT is not expected to be common among or play the same role in the evolution of multicellular plants and animals, especially those with sequestered germ lines, and there simply is no extensive data on LGT in unicellular eukaryotes.)

What If Phylogenetic Classification Is Just Let Go?

Before Darwin, the purpose of a natural classification was to reveal divine or other eternal ordering principles that explained patterns of similarity and difference between species. Darwin argued that "propinquity of descent—the only known cause of the similarity of organic beings—is the bond, hidden as it is by various degrees of modification, which is partially revealed to us by our classification" (*3*, p. 399). Biologists came to think that living species diverge from an ever smaller number of ancestral species, back to the very first organism, such that the ultimate natural order is a single inclusively hierarchical, "universal phylogenetic tree," without reticulation. They might not be certain as to which genus (or kingdom) a species (or phylum) belonged, but there was no

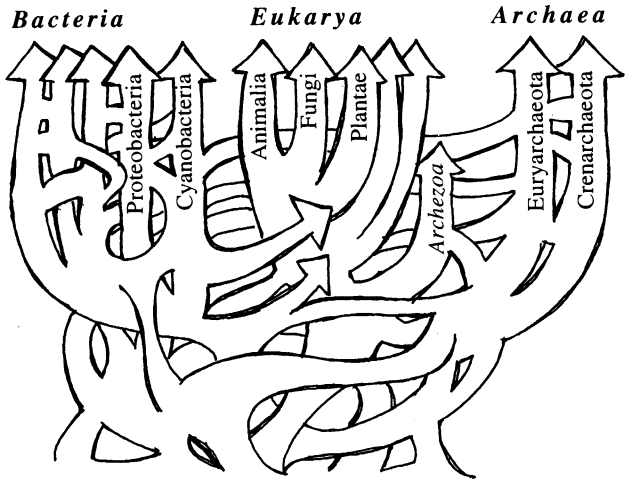

Fig. 3. A reticulated tree, or net, which might more appropriately represent life's history. Martin (*16*), in a review covering many of the same topics as this one, has presented some striking colored representations of such patterns.

question of its belonging to more than one.

After Zuckerkandl and Pauling (*4*), biologists came to think that the universal tree could be reduced to a tree based on sequences of orthologous genes, any of which (practical considerations aside) could serve as a marker for an entire genome, organism, or species. If, however, different genes give different trees, and there is no fair way to suppress this disagreement, then a species (or phylum) can "belong" to many genera (or kingdoms) at the same time: There really can be no universal phylogenetic tree of organisms based on such a reduction to genes.

To save the trees, one might define organisms as more than the sums of their genes and imagine organismal lineages to have a sort of emergent reality—just as we think of ourselves as real and continuous over a lifetime, while knowing that we contain very few of the atoms with which we were born. But, one cannot learn about the histories of such emergent entities by studying the histories of their individual parts unless arguable assumptions of the sort discussed above are made.

For prokaryotes, LGT compromises the definition of taxa at all ranks, especially the highest. Archaea (or Bacteria) may well be definable by sets of genes conserved within and not between them, but the hierarchical pattern shown in Fig. 2 is only one of many possible depictions of relationships between individual archaeal or bacterial genes and is thus not a fair (at least not complete) depiction of the actual evolutionary history of any lineage of real organisms.

Perhaps it would be easier, and in the long run more productive, to abandon the attempt to force the data that Zuckerkandl and Pauling stimulated biologists to collect into the mold provided by Darwin. If there were believable genealogies of all genes (and intragenic recombination could be ignored), one could then ask which genes have traveled together for how long in which genomes, without an obligation to marshal these data in the defense of one or another grander phylogenetic scheme for organisms. One could, as Martin (*16*, p. 104) has exhorted, set about discovering the "principles which must govern the distribution of genes across bacterial genomes." While retaining useful names for recognized groups (Archaea and Bacteria), one could see these as taxonomic descriptors based directly on shared genes, but only based indirectly and unpredictably on shared ancestry. As an example, one could then easily accept the fact that the cyanobacteria appear to be a very "good" taxon in the sense that many molecules support their monophyly but that they nevertheless might derive major elements of their uniquely defining photosynthetic biochemistry from different bacterial antecedents (*42*). As another, one might cease being surprised or upset that the obvious extensive sharing of genes between thermophilic prokaryotes makes the placement of any individual thermophile in its "true" position in the tree a highly problematic exercise (*43*). In other words, biologists might rejoice in and explore, rather than regret or attempt to dismiss, the creative evolutionary role of LGT.

References and Notes

1. E. Mayr, *The Growth of Biological Thought* (Belknap Press, Cambridge, MA, 1982).
2. A. L. Panchen, *Classification, Evolution and the Nature of Biology* (Cambridge Univ. Press, Cambridge, 1992); M. T. Ghiselin, *Metaphysics and the Origin of Species* (State Univ. of New York Press, Albany, NY, 1997).
3. C. Darwin, *The Origin of Species by Means of Natural Selection* (Murray, London, 1859).
4. E. Zuckerkandl and L. Pauling, in *Evolving Genes and Proteins*, V. Bryson and H. J. Vogel, Eds. (Academic Press, New York, 1965), pp. 97–166; *J. Theor. Biol.* **8**, 357 (1965).
5. C. R. Woese, O. Kandler, M. L. Wheelis, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576 (1990); W. F. Doolittle and J. R. Brown, *ibid.* **91**, 6721 (1994); R. F. Doolittle, *ibid.* **92**, 2421 (1995); N. R. Pace, *Science* **276**, 734 (1997).
6. R. M. Schwartz and M. O. Dayhoff, *Science* **199**, 395 (1978).
7. C. R. Woese, *Microbiol. Rev.* **51**, 221 (1987).
8. R. Green and H. F. Noller, *Annu. Rev. Biochem.* **66**, 679 (1997).
9. P. J. Gogarten *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6661 (1989); N. Iwabe, K. Kuma, M. Hasegawa, S. Osawa, T. Miyata, *ibid.*, p. 9355; J. R. Brown and W. F. Doolittle, *ibid.* **92**, 2441 (1995); S. L. Baldauf, J. D. Palmer, W. F. Doolittle, *ibid.* **93**, 7749 (1996); F. S. Lawson, R. L. Charlebois, J. A. Dillon, *Mol. Biol. Evol.* **13**, 970 (1996). Organisms in all major groups bear genes encoding elongation factors EF-1α (EF-Tu) and EF-2 (bacterial EF-G). These homologous (more precisely, paralogous) genes thus must be products of a gene duplication that had already occurred before the time of the last common ancestor of all organisms alive today. A tree made of EF-1α and EF-2 sequences should comprise two subtrees (one for each of these paralogs), each of which (barring LGT or artifacts) should have the topology of the true universal organismal tree. The point at which each subtree attaches to the other will be its root and will correspond to the last common organismal ancestor. Neither subtree is rootable without the other. Because rRNA genes all seem to be orthologous (derived from a single ancestral gene in the last common ancestor), such a rooting is not possible with rRNA sequences alone.
10. M. J. Kates, D. J. Kushner, A. T. Matheson, Eds., *The Biochemistry of Archaea (Archaeobacteria)* (Elsevier Science, Amsterdam, 1993); W. F. Doolittle, in *Evolution of Microbial Life*, D. M. Roberts, P. Sharp, G. Alderson, M. Collins, Eds. (Cambridge Univ. Press, Cambridge, 1996), pp. 1–21.
11. G. J. Olsen and C. R. Woese, *Cell* **89**, 991 (1997); D. E. Edgell and W. F. Doolittle, *ibid.*, p. 995; J. N. Reeve, K. Sandman, C. J. Daniels, *ibid.*, p. 999; P. P. Dennis, *ibid.*, p. 1007; J. Soppa, *Mol. Microbiol.* **31**, 1295 (1997).
12. H. Philippe, *Curr. Opin. Genet. Dev.* **8**, 616 (1998); H. Phillippe and A. Adoutte, in *Evolutionary Relationships Among Protozoa*, G. H. Coombs, K. Vickerman, M. A. Sleigh, A. Warren, Eds. (Systematics Association, London, 1998), pp. 25–26; D. Moreira, H. LeGuyader, H. Philippe, *Mol. Biol. Evol.* **16**, 234 (1999).
13. Prominent among sources of error or uncertainty in establishing branching patterns are mutational saturation, "long-branch attraction," and "among-site rate variation." Mutationally saturated sequences are maximally diverged, so that further changes are as likely to make them more similar as they are to make them more different, and tree topology is based on noise. Long-branch attraction [J. Felsenstein, *Syst. Zool.* **27**, 401 (1978)] occurs when rates of sequence change differ substantially between taxa (even without saturation). Lineages with higher rates of sequence change artifactually associate with each other and with outgroups, except with maximum likelihood methods. Even with these methods [Z. Yang, *J. Mol. Evol.* **42**, 294 (1996); *Trends Ecol. Evol.* **11**, 367 (1996)], long-branch attraction occurs when there is a substantial rate variation among different sites in a gene.
14. R. P. Hirt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 580 (1999); J. W. Stiller, E. C. Duffield, B. D. Hall, *ibid.* **95**, 11769 (1998).
15. P. J. Gogarten, E. Hilario, L. Olenzenski, in *Evolution of Microbial Life*, D. M. Roberts, P. Sharp, G. Alderson, M. Collins, Eds. (Cambridge Univ. Press, Cambridge, 1996), pp. 267–292.
16. W. Martin, *BioEssays* **21**, 99 (1999).
17. T. Cavalier-Smith, *Nature* **326**, 332 (1987); *Biol. Rev. Camb. Philos. Soc.* **73**, 203 (1998).
18. C. R. Vossbrinck *et al.*, *Nature* **326**, 411 (1987); M. L. Sogin, J. H. Gunderson, H. J. Elwood, R. A. Alonso, D. A. Peattie, *Science* **243**, 75 (1989); D. D. Leipe, J. H. Gunderson, T. A. Nerad, M. L. Sogin, *Mol. Biochem. Parasitol.* **59**, 41 (1993).
19. T. M. Embley and R. P. Hirt, *Curr. Opin. Genet. Dev.* **8**, 629 (1998); A. J. Roger, O. Sandblom, W. F. Doolittle, H. Philippe, *Mol. Biol. Evol.* **16**, 218 (1999); P. J. Keeling and G. I. McFadden, *Trends Microbiol.* **6**, 19 (1998).
20. M. M. Miyamoto and W. M. Fitch, *Mol. Biol. Evol.* **12**, 513 (1995).
21. M. W. Gray, G. Burger, B. F. Lang, *Science* **283**, 1476 (1999).
22. J. R. Brown and W. F. Doolittle, *Microbiol. Mol. Biol. Rev.* **61**, 456 (1997); D.-F. Feng, G. Cho, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13028 (1997); M. Ragan and T. Gaasterland, *J. Microb. Comp. Genomics* **3**, 219 (1998); M. C. Rivera, R. Jain, F. F. Moore, J. A. Lake, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6239 (1998).
23. W. F. Doolittle, *Trends Genet.* **14**, 307 (1998); W. Martin and M. Muller, *Nature* **392**, 37 (1998).
24. J. G. Lawrence and H. Ochman, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9413 (1998).
25. H.-P. Klenk *et al.*, *Nature* **390**, 364 (1997).
26. W. F. Doolittle and J. M. Logsdon Jr., *Curr. Biol.* **8**, R209 (1998); M. Ibba, J. L. Bono, P. A. Rosa, D. Soll, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14383 (1997).
27. R. Jain, M. C. Rivera, J. A. Lake, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3801 (1999).
28. E. Hilario and J. P. Gogarten, *Biosystems* **31**, 111 (1993).
29. S. Sonea and M. Paniset, *Rev. Can. Biol.* **35**, 103 (1976); D. C. Reanney, in *Aspects of Genetic Action and Evolution*, suppl. 8 of *International Review of Cytology*, G. H. Bourne, J. F. Danielli, K. W. Jeon, Eds. (Academic Press, New York, 1978), pp. 1–67.
30. N. Cermakian *et al.*, *J. Mol. Evol.* **46**, 671 (1997).
31. S. A. Baldauf and W. F. Doolittle, in preparation.
32. D. M. O'Neil, L. Baron, P. Sypherd, *J. Bacteriol.* **99**, 242 (1969).
33. R. F. Doolittle and J. F. Handy, *Curr. Opin. Genet. Dev.* **8**, 630 (1998).
34. M. Nomura, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1820 (1999).
35. _____, P. Traub, H. Bechmann, *Nature* **219**, 793 (1968).
36. H. F. Noller and C. R. Woese, *Science* **212**, 403 (1981).
37. T. Asai, D. Zaprojets, C. Squires, C. L. Squires, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1971 (1999).
38. R. S. Gupta, *Microbiol. Mol. Biol. Rev.* **62**, 1435 (1998).
39. J. R. Brown, J. Zhang, J. E. Hodgson, *Curr. Biol.* **8**, R365 (1998).
40. K. Ueda, Y. Kido, T. Yoshida, M. Kataoka, *J. Bacteriol.* **181**, 78 (1999).
41. C. R. Woese, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6854 (1998).
42. J. Xiong, K. Inoue, C. C. Bauer, *ibid.*, p. 14851.
43. L. Aravind *et al.*, *Trends Genet.* **14**, 442 (1998).

44. I thank J. Logsdon, A. Roger, D. Faguy, O. Feeley, and Y. Inagaki for critical discussions and the Medical Research Council of Canada and the Canadian Institute for Advanced Research for support. I am indebted to J. P. Gogarten and W. Martin for persuading me of the importance of LGT.