Uprooting the Tree of Life

About 10 years ago scientists finally worked out the basic outline of how modern life-forms evolved. Now parts of their tidy scheme are unraveling

by W. Ford Doolittle

Charles Darwin contended more than a century ago that all modern species diverged from a more limited set of ancestral groups, which themselves evolved from still fewer progenitors and so on back to the beginning of life. In principle, then, the relationships among all living and extinct organisms could be represented as a single genealogical tree.

Most contemporary researchers agree. Many would even argue that the general features of this tree are already known, all the way down to the root—a solitary cell, termed life's last universal common ancestor, that lived roughly 3.5 to 3.8 billion years ago. The consensus view did not come easily but has been widely accepted for more than a decade.

Yet ill winds are blowing. To everyone's surprise, discoveries made in the past few years have begun to cast serious doubt on some aspects of the tree, especially on the depiction of the relationships near the root.

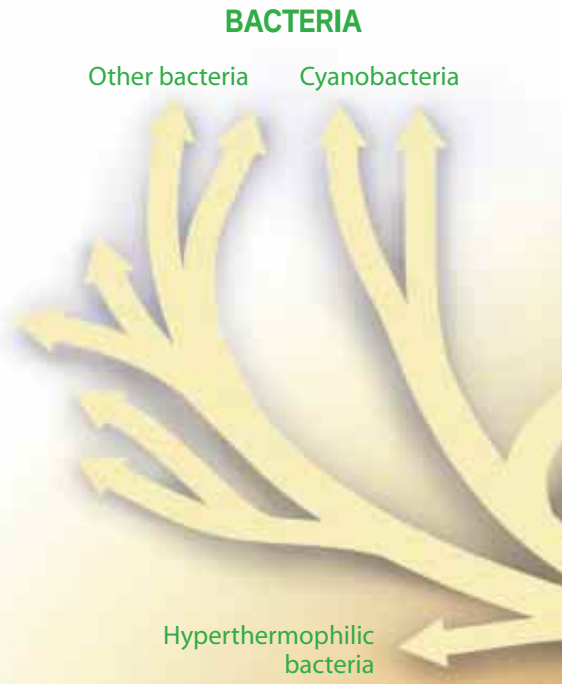
The First Sketches

Scientists could not even begin to contemplate constructing a universal tree until about 35 years ago. From the time of Aristotle to the 1960s, researchers deduced the relatedness of organisms by comparing their anatomy or physiology, or both. For complex organisms, they were frequently able to draw reasonable genealogical inferences in this way. Detailed analyses of innumerable traits suggested, for instance, that hominids shared a common ancestor with apes, that this common ancestor shared an earlier one with monkeys, and that *that* precursor shared an even earlier forebear with prosimians, and so forth.

Microscopic single-celled organisms, however, often provided too little information for defining relationships. That paucity was disturbing because microbes were the only inhabitants of the earth for the first half to two thirds of the planet's history; the absence of a clear phylogeny (family tree) for microorganisms left scientists unsure about the sequence in which some of the most radical innovations in cellular structure and function occurred. For example, between the birth of the first cell and the appearance of multicellular fungi, plants and animals, cells grew bigger and more complex, gained a nucleus and a cytoskeleton (internal scaffolding), and found a way to eat other cells.

In the mid-1960s Emile Zuckerkandl and Linus Pauling of the California Institute of Technology conceived of a revolutionary strategy that could supply the missing information. Instead of looking just at anatomy or physiology, they asked, why not base family trees on differences in the order of the building blocks in selected genes or proteins?

Their approach, known as molecular phylogeny, is eminently logical. Individual genes, composed of unique sequences of nucleotides, typically serve as the blueprints for making specific proteins, which consist of particular strings of amino acids. All genes, however, mutate (change in sequence), sometimes altering the encoded protein. Genetic mutations that have no effect on protein function or that improve it will inevitably accumulate over time. Thus, as two species diverge from an ancestor, the sequences of the genes they share will also diverge. And as time passes, the genetic divergence will increase. Investigators can therefore recon-



CONSENSUS VIEW of the universal tree of life holds that the early descendants of life's last universal common ancestor—a small cell with no nucleus—divided into two prokaryotic (nonnucleated) groups: the bacteria and the archaea. Later, the archaea gave rise to organisms having complex cells containing a nucleus: the eukaryotes. Eukaryotes gained valuable energy-generating organelles—mitochondria and, in the case of plants, chloroplasts—by taking up, and retaining, certain bacteria.

Copyright 2000 Scientific American, Inc.

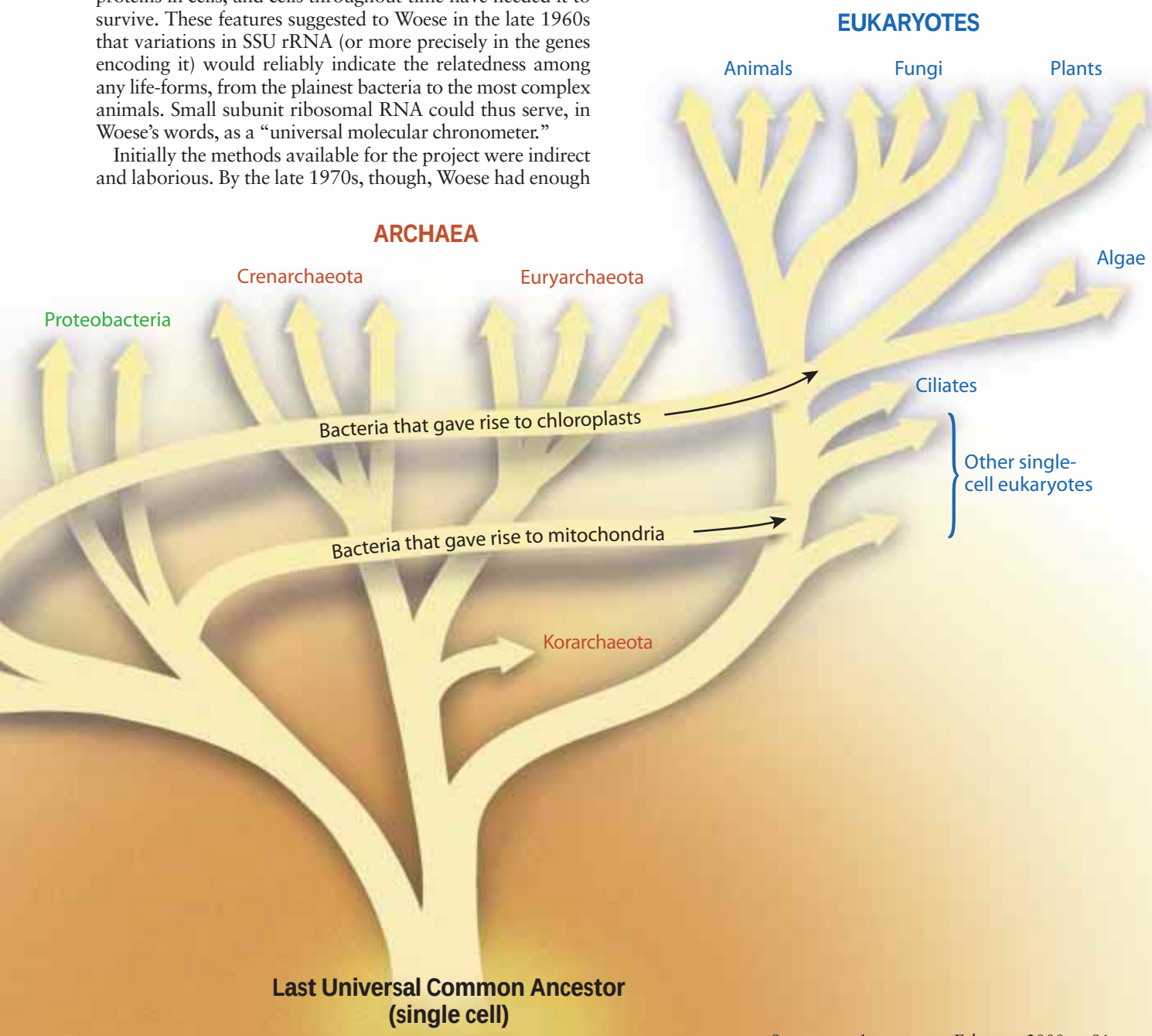
struct the evolutionary past of living species—can construct their phylogenetic trees—by assessing the sequence divergence of genes or proteins isolated from those organisms.

Thirty-five years ago scientists were just becoming proficient at identifying the order of amino acids in proteins and could not yet sequence genes. Protein studies completed in the 1960s and 1970s demonstrated the general utility of molecular phylogeny by confirming and then extending the family trees of well-studied groups such as the vertebrates. They also lent support to some hypotheses about the links among certain bacteria—showing, for instance, that bacteria capable of producing oxygen during photosynthesis form a group of their own (cyanobacteria).

As this protein work was progressing, Carl R. Woese of the University of Illinois was turning his attention to a powerful new yardstick of evolutionary distances: small subunit ribosomal RNA (SSU rRNA). This genetically specified molecule is a key constituent of ribosomes, the "factories" that construct proteins in cells, and cells throughout time have needed it to survive. These features suggested to Woese in the late 1960s that variations in SSU rRNA (or more precisely in the genes encoding it) would reliably indicate the relatedness among any life-forms, from the plainest bacteria to the most complex animals. Small subunit ribosomal RNA could thus serve, in Woese's words, as a "universal molecular chronometer."

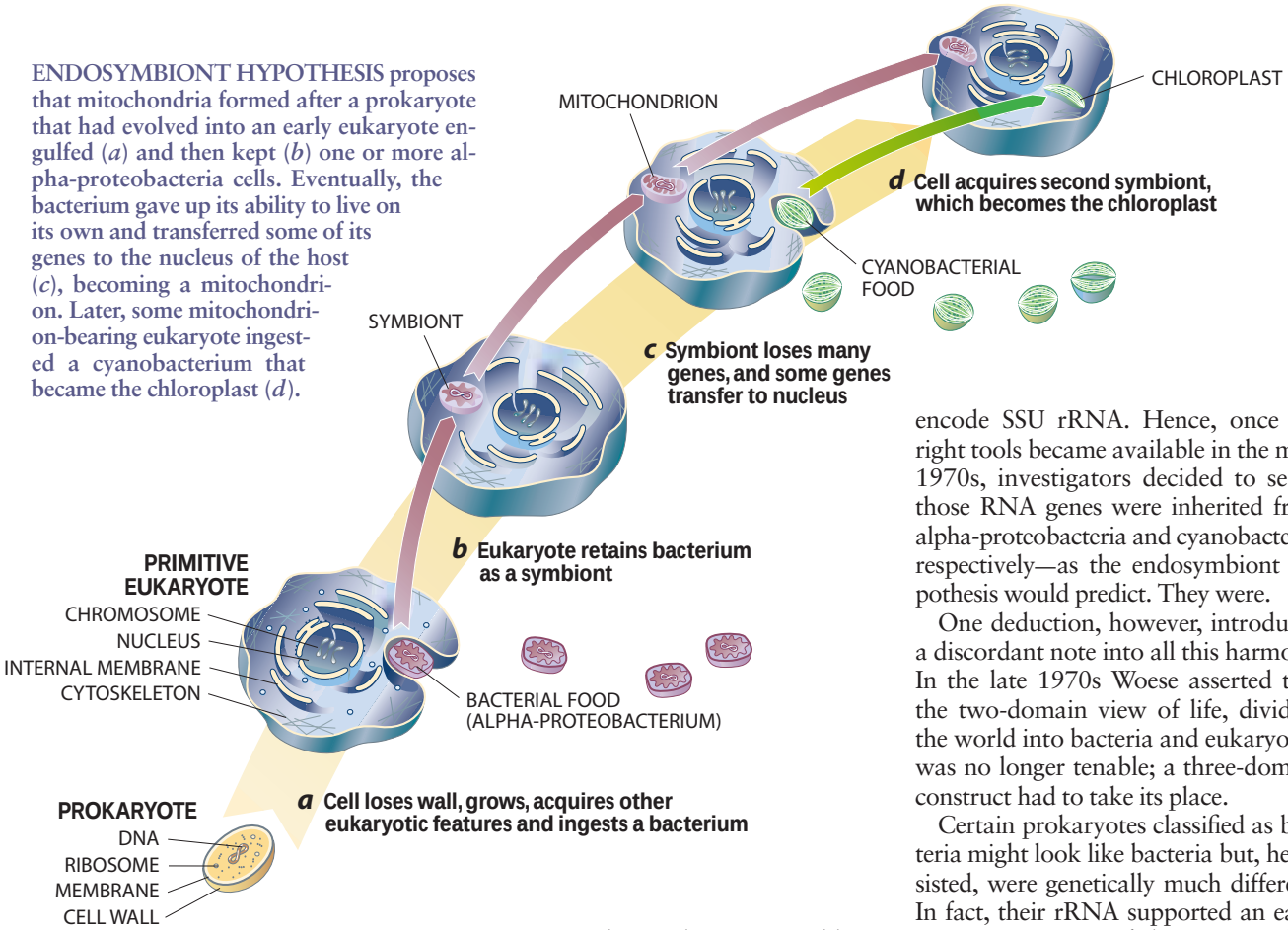
Initially the methods available for the project were indirect and laborious. By the late 1970s, though, Woese had enough data to draw some important inferences. Since then, phylogeneticists studying microbial evolution, as well as investigators concerned with higher sections of the universal tree, have based many of their branching patterns on sequence analyses of SSU rRNA genes. This accumulation of rRNA data helped greatly to foster consensus about the universal tree in the late 1980s. Today investigators have rRNA sequences for several thousands of species.

From the start, the rRNA results corroborated some already accepted ideas, but they also produced an astonishing surprise. By the 1960s microscopists had determined that the world of living things could be divided into two separate groups, eukaryotes and prokaryotes, depending on the structure of the cells that composed them. Eukaryotic organisms (animals, plants, fungi and many unicellular life-forms) were defined as those composed of cells that contained a true nucleus—a membrane-bound organelle housing the chromosomes. Eukaryotic cells



JANA BRENNING

Copyright 2000 Scientific American, Inc.



ENDOSYMBIONT HYPOTHESIS proposes that mitochondria formed after a prokaryote that had evolved into an early eukaryote engulfed (*a*) and then kept (*b*) one or more alpha-proteobacteria cells. Eventually, the bacterium gave up its ability to live on its own and transferred some of its genes to the nucleus of the host (*c*), becoming a mitochondrion. Later, some mitochondrion-bearing eukaryote ingested a cyanobacterium that became the chloroplast (*d*).

CHRISTOPH BLUMRICH

also displayed other prominent features, among them a cytoskeleton, an intricate system of internal membranes and, usually, mitochondria (organelles that perform respiration, using oxygen to extract energy from nutrients). In the case of algae and higher plants, the cells also contained chloroplasts (photosynthetic organelles).

Prokaryotes, thought at the time to be synonymous with bacteria, were noted to consist of smaller and simpler nonnucleated cells. They are usually enclosed by both a membrane and a rigid outer wall.

Woese's early data supported the distinction between prokaryotes and eukaryotes, by establishing that the SSU rRNAs in typical bacteria were more similar in sequence to one another than to the rRNA of eukaryotes. The initial rRNA findings also lent credence to one of the most interesting notions in evolutionary cell biology: the endosymbiont hypothesis. This conception aims to explain how eukaryotic cells first came to possess mitochondria and chloroplasts [see "The Birth of Complex Cells," by Christian de Duve, Scientific American, April 1996].

On the way to becoming a eukaryote, the hypothesis proposes, some ancient anaerobic prokaryote (unable to use oxygen for energy) lost its cell wall. The more flexible membrane underneath then began to grow and fold in on itself. This change, in turn, led to formation of a nucleus and other internal membranes and also enabled the cell to engulf and digest neighboring prokaryotes, instead of gaining nourishment entirely by absorbing small molecules from its environment.

At some point, one of the descendants of this primitive eukaryote took up bacterial cells of the type known as alpha-proteobacteria, which are proficient at respiration. But instead of digesting this "food," the eukaryote settled into a mutually beneficial (symbiotic) relationship with it. The eukaryote sheltered the internalized cells, and the "endosymbionts" provided extra energy to the host through respiration. Finally, the endosymbionts lost the genes they formerly used for independent growth and transferred others to the host's nucleus—becoming mitochondria in the process. Likewise, chloroplasts derive from cyanobacteria that an early, mitochondria-bearing eukaryote took up and kept.

Mitochondria and chloroplasts in modern eukaryotes still retain a small number of genes, including those that encode SSU rRNA. Hence, once the right tools became available in the mid-1970s, investigators decided to see if those RNA genes were inherited from alpha-proteobacteria and cyanobacteria, respectively—as the endosymbiont hypothesis would predict. They were.

One deduction, however, introduced a discordant note into all this harmony. In the late 1970s Woese asserted that the two-domain view of life, dividing the world into bacteria and eukaryotes, was no longer tenable; a three-domain construct had to take its place.

Certain prokaryotes classified as bacteria might look like bacteria but, he insisted, were genetically much different. In fact, their rRNA supported an early separation. Many of these species had already been noted for displaying unusual behavior, such as favoring extreme environments, but no one had disputed their status as bacteria. Now Woese claimed that they formed a third primary group—the archaea—as different from bacteria as bacteria are from eukaryotes.

Acrimony, Then Consensus

At first, the claim met enormous resistance. Yet eventually most scientists became convinced, in part because the overall structures of certain molecules in archaeal species corroborated the three-group arrangement. For instance, the cell membranes of all archaea are made up of unique lipids (fatty substances) that are quite distinct—in their physical properties, chemical constituents and linkages—from the lipids of bacteria.

Similarly, the archaeal proteins responsible for several crucial cellular processes have a distinct structure from the proteins that perform the same tasks in bacteria. Gene transcription and translation are two of those processes. To make a protein, a cell first copies, or transcribes, the corresponding gene into a strand of messenger RNA. Then ribosomes translate the messenger RNA codes into a specific string of amino acids. Bio-

Copyright 2000 Scientific American, Inc.

chemists found that archaeal RNA polymerase, the enzyme that carries out gene transcription, more resembles its eukaryotic than its bacterial counterparts in complexity and in the nature of its interactions with DNA. The protein components of the ribosomes that translate archaeal messenger RNAs are also more like the ones in eukaryotes than those in bacteria.

Once scientists accepted the idea of three domains of life instead of two, they naturally wanted to know which of the two structurally primitive groups—bacteria or archaea—gave rise to the first eukaryotic cell. The studies that showed a kinship between the transcription and translation machinery in archaea and eukaryotes implied that eukaryotes diverged from the archaeans.

This deduction gained added credibility in 1989, when groups led by J. Peter Gogarten of the University of Connecticut and Takashi Miyata, then at Kyushu University in Japan, used sequence information from genes for other cellular components to "root" the universal tree. Comparisons of SSU rRNA can indicate which organisms are closely related to one another but, for technical reasons, cannot by themselves indicate which groups are oldest and therefore closest to the root of the tree. The DNA sequences encoding two essential cellular proteins agreed that the last common ancestor spawned both the bacteria and the archaea; then the eukaryotes branched from the archaea.

Since 1989 a host of discoveries have supported that depiction. In the past five years, sequences of the full genome (the total complement of genes) in half a dozen archaea and more than 15 bacteria have become available. Comparisons of such genomes confirm earlier suggestions that many genes involved in transcription and translation are much the same in eukaryotes and archaea and that these processes are performed very similarly in the two domains. Further, although archaea do not have nuclei, under certain experimental conditions their chromosomes resemble those of eukaryotes: the DNA appears to be associated with eukaryote-type proteins called histones, and the chromosomes can adopt a eukaryotic "beads-on-a-string" structure. These chromosomes are replicated by a suite of proteins, most of which are found in some form in eukaryotes but not in bacteria.

Nevertheless, Doubts

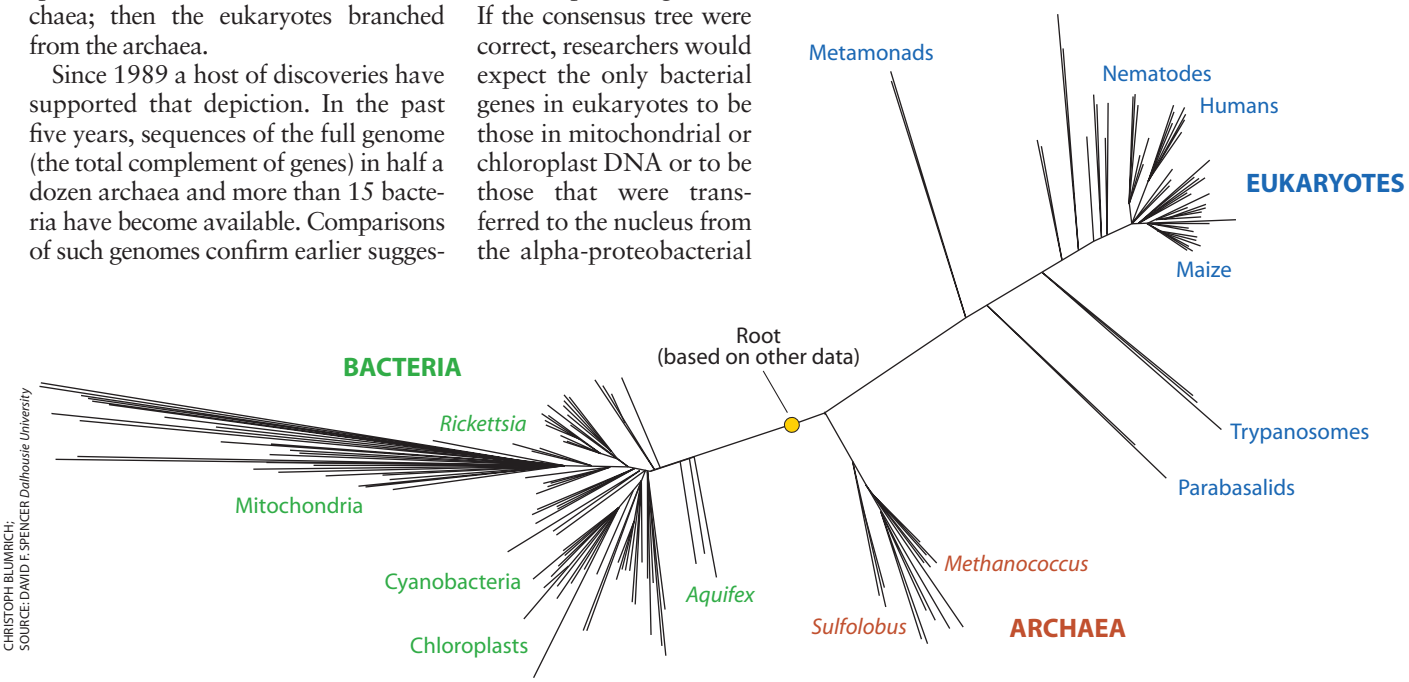
The accumulation of all these wonderfully consistent data was gratifying and gave rise to the now accepted arrangement of the universal genealogical tree. This phylogeny indicates that life diverged first into bacteria and archaea. Eukaryotes then evolved from an archaealike precursor. Subsequently, eukaryotes took up genes from bacteria twice, obtaining mitochondria from alpha-proteobacteria and chloroplasts from cyanobacteria.

Still, as DNA sequences of complete genomes have become increasingly available, my group and others have noted patterns that are disturbingly at odds with the prevailing beliefs. If the consensus tree were correct, researchers would expect the only bacterial genes in eukaryotes to be those in mitochondrial or chloroplast DNA or to be those that were transferred to the nucleus from the alpha-proteobacterial or cyanobacterial precursors of these organelles. The transferred genes, moreover, would be ones involved in respiration or photosynthesis, not in cellular processes that would already be handled by genes inherited from the ancestral archaean.

Those expectations have been violated. Nuclear genes in eukaryotes often derive from bacteria, not solely from archaea. A good number of those bacterial genes serve nonrespiratory and nonphotosynthetic processes that are arguably as critical to cell survival as are transcription and translation.

The classic tree also indicates that bacterial genes migrated only to a eukaryote, not to any archaea. Yet we are seeing signs that many archaea possess a substantial store of bacterial genes. One example among many is *Archaeoglobus fulgidus*. This organism meets all the criteria for an archaean (it has all the proper lipids in its cell membrane and the right transcriptional and translational machinery), but it uses a bacterial form of the enzyme HMGCoA reductase for synthesizing membrane lipids. It also has numerous bacterial genes that help it to gain energy and nutrients in one of its favorite habitats: undersea oil wells.

The most reasonable explanation for these various contrarian results is that the pattern of evolution is not as linear and treelike as Darwin imagined it. Al-



CHRISTOPH BLUMRICH; SOURCE: DAVID F. SPENCER *Dalhousie University*

RELATIONSHIPS among ribosomal RNAs (rRNAs) from almost 600 species are depicted. A single line represents the rRNA sequence in one species or a group; many of the lines reflect rRNAs encoded by nuclear genes, but others reflect rRNAs encoded by chloroplast or mitochondrial genes. The mitochondrial lines are relatively long because mitochondrial genes evolve rapidly. Trees derived from rRNA data are rootless; other data put the root at the colored dot, corresponding to the lowest part of the tree on pages 90 and 91.

Copyright 2000 Scientific American, Inc.

though genes are passed vertically from generation to generation, this vertical inheritance is not the only important process that has affected the evolution of cells. Rampant operation of a different process—lateral, or horizontal, gene transfer—has also affected the course of that evolution profoundly. Such transfer involves the delivery of single genes, or whole suites of them, not from a parent cell to its offspring but across species barriers.

Lateral gene transfer would explain how eukaryotes that supposedly evolved from an archaeal cell obtained so many bacterial genes important to metabolism: the eukaryotes picked up the genes from bacteria and kept those that proved useful. It would likewise explain how various archaea came to possess genes usually found in bacteria.

Some molecular phylogenetic theorists—among them, Mitchell L. Sogin of the Marine Biological Laboratory in Woods Hole, Mass., and Russell F. Doolittle (my very distant relative) of the University of California at San Diego—have also invoked lateral gene transfer to explain a long-standing mystery. Many eukaryotic genes turn out to be unlike those of any known archaea or bacteria; they seem to have come from nowhere. Notable in this regard are the genes for the components of two defining eukaryotic features, the cytoskeleton and the system of internal membranes. Sogin and Doolittle suppose that some fourth domain of organisms, now extinct, slipped those surprising genes into the eukaryotic nuclear genome horizontally.

In truth, microbiologists have long known that bacteria exchange genes horizontally. Gene swapping is clearly how some disease-causing bacteria give the gift of antibiotic resistance to other species of infectious bacteria. But few researchers suspected that genes essential to the very survival of cells traded hands frequently or that lateral transfer exerted great influence on the early history of microbial life. Apparently, we were mistaken.

Can the Tree Survive?

What do the new findings say about the structure of the universal tree of life? One lesson is that the neat progression from archaea to eukaryote in the consensus tree is oversimplified or wrong. Plausibly, eukaryotes emerged not from an archaean but from some precursor cell that was the product of any number of horizontal gene transfers—events that left it part bacterial and part archaean and maybe part other things.

The weight of evidence still supports the likelihood that mitochondria in eukaryotes derived from alpha-proteobacterial cells and that chloroplasts came from ingested cyanobacteria, but it is no longer safe to assume that those were the only lateral gene transfers that occurred after the first eukaryotes arose. Only in later, multicellular eukaryotes do we know of definite restrictions on horizontal gene exchange, such as the advent of separated (and protected) germ cells.

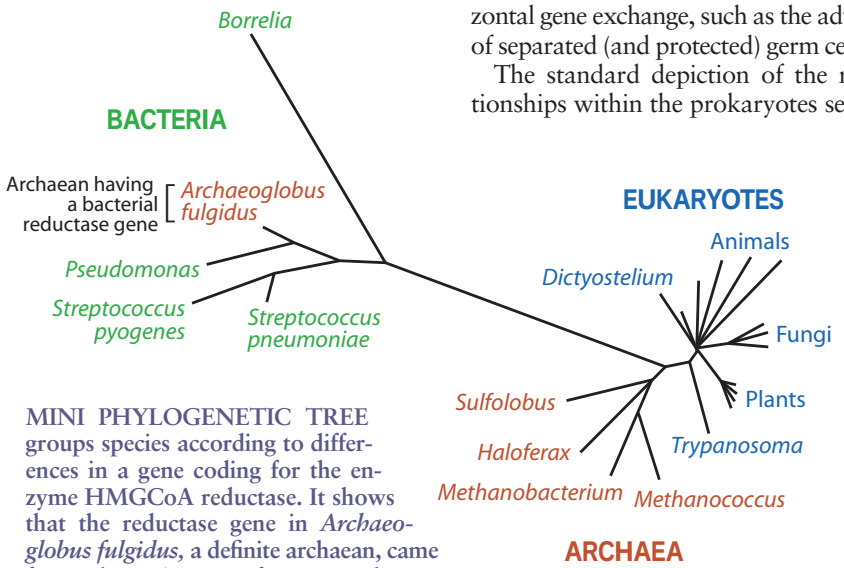
The standard depiction of the relationships within the prokaryotes seems too pat as well. A host of genes and biochemical features do unite the prokaryotes that biologists now call archaea and distinguish those organisms from the prokaryotes we call bacteria, but bacteria and archaea (as well as species within each group) have clearly engaged in extensive gene swapping.

Researchers might choose to define evolutionary relationships within the prokaryotes on the basis of genes that seem least likely to be transferred. Indeed, many investigators still assume that genes for SSU rRNA and the proteins involved in transcription and translation are unlikely to be moveable and that the phylogenetic tree based on them thus remains valid. But this nontransferability is largely an untested assumption, and in any case, we must now admit that any tree is at best a description of the evolutionary history of only part of an organism's genome. The consensus tree is an overly simplified depiction.

What would a truer model look like? At the top, treelike branching would continue to be apt [*see illustration on opposite page*] for multicellular animals, plants and fungi. And gene transfers involved in the formation of bacteria-derived mitochondria and chloroplasts in eukaryotes would still appear as fusions of major branches. Below these transfer points (and continuing up into the modern bacterial and archaeal domains), we would, however, see a great many additional branch fusions. Deep in the realm of the prokaryotes and perhaps at the base of the eukaryotic domain, designation of any trunk as the main one would be arbitrary.

Though complicated, even this revised picture would actually be misleadingly simple, a sort of shorthand cartoon, because the fusing of branches usually would not represent the joining of whole genomes, only the transfers of single or multiple genes. The full picture would have to display simultaneously the superimposed genealogical patterns of thousands of different families of genes (the rRNA genes form just one such family).

If there had never been any lateral transfer, all these individual gene trees would have the same topology (the same branching order), and the ancestral genes at the root of each tree would have all been present in the genome of the universal last common ancestor, a single ancient cell. But extensive transfer means that neither is the case: gene trees will differ (although many will have regions of similar topology), *and*



MINI PHYLOGENETIC TREE groups species according to differences in a gene coding for the enzyme HMGCoA reductase. It shows that the reductase gene in *Archaeoglobus fulgidus,* a definite archaean, came from a bacterium, not from an archaean ancestor. This finding is part of growing evidence indicating that the evolution of unicellular life has long been influenced profoundly by lateral gene transfer (occurring between contemporaries). The consensus universal tree does not take that influence into account.

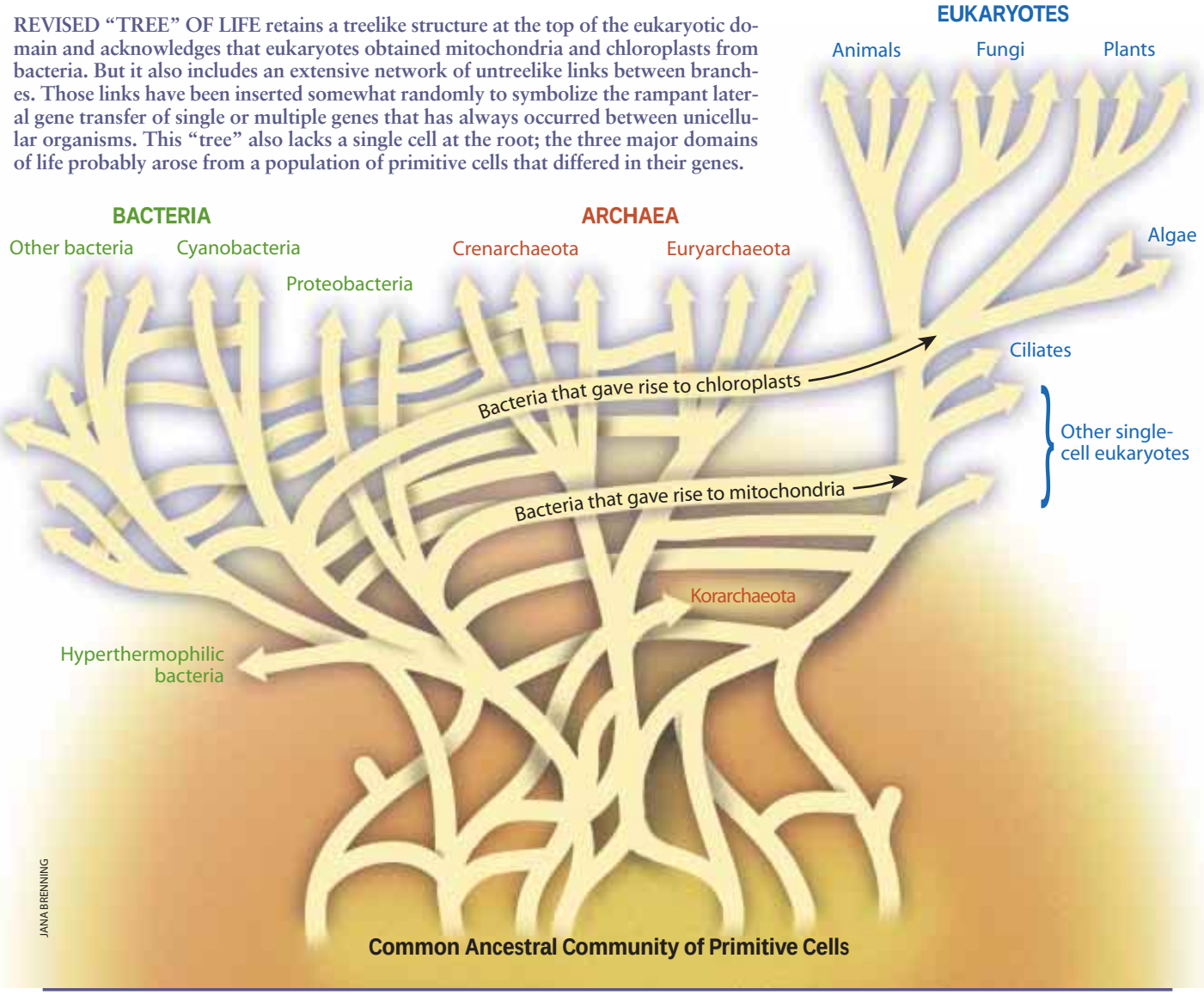
CHRISTOPH BLUMRICH

Copyright 2000 Scientific American, Inc.

there would never have been a single cell that could be called the last universal common ancestor.

As Woese has written, "The ancestor cannot have been a particular organism, a single organismal lineage. It was communal, a loosely knit, diverse conglomeration of primitive cells that evolved as a unit, and it eventually developed to a stage where it broke into several distinct communities, which in their turn become the three primary lines of descent [bacteria, archaea and eukaryotes]." In other words, early cells, each having relatively few genes, differed in many ways. By swapping genes freely, they shared various of their talents with their contemporaries. Eventually this collection of eclectic and changeable cells coalesced into the three basic domains known today. These domains remain recognizable because much (though by no means all) of the gene transfer that occurs these days goes on within domains.

Some biologists find these notions confusing and discouraging. It is as if we have failed at the task that Darwin set for us: delineating the unique structure of the tree of life. But in fact, our science is working just as it should. An attractive hypothesis or model (the single tree) suggested experiments, in this case the collection of gene sequences and their analysis with the methods of molecular phylogeny. The data show the model to be too simple. Now new hypotheses, having final forms we cannot yet guess, are called for. SA



REVISED "TREE" OF LIFE retains a treelike structure at the top of the eukaryotic domain and acknowledges that eukaryotes obtained mitochondria and chloroplasts from bacteria. But it also includes an extensive network of untreelike links between branches. Those links have been inserted somewhat randomly to symbolize the rampant lateral gene transfer of single or multiple genes that has always occurred between unicellular organisms. This "tree" also lacks a single cell at the root; the three major domains of life probably arose from a population of primitive cells that differed in their genes.

The Author

W. FORD DOOLITTLE, who holds degrees from Harvard and Stanford universities, is professor of biochemistry and molecular biology at Dalhousie University in Halifax, Nova Scotia, and director of the Program in Evolutionary Biology of the Canadian Institute for Advanced Research.

Further Information

THE UNIVERSAL ANCESTOR. Carl Woese in the *Proceedings of the National Academy of Sciences,* Vol. 95, No. 12, pages 6854–6859; June 9, 1998.

YOU ARE WHAT YOU EAT: A GENE TRANSFER RACHET COULD ACCOUNT FOR BACTERIAL GENES IN EUKARYOTIC NUCLEAR GENOMES. W. Ford Doolittle in *Trends in Genetics,* Vol. 14, No. 8, pages 307–311; August 1998.

PHYLOGENETIC CLASSIFICATION AND THE UNIVERSAL TREE. W. Ford Doolittle in *Science,* Vol. 284, pages 2124–2128; June 25, 1999.

Copyright 2000 Scientific American, Inc.