2 What Is a Prokaryote?

W. Ford Doolittle[1] · *Olga Zhaxybayeva*[2]
[1]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada
[2]Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire, USA

Some Tangled-Together Questions About "Prokaryote" ... 21

History and Fact ... 23

The Concept of a Bacterium ... 24

Three Domains Versus Two: The Debate with Ernst Mayr ... 26

Time for a Change? ... 27

Paraphyly ... 28

"Pro" Implies Before ... 28

Negative Definition ... 28

Bacteria and Archaea, and the Fundamental Differences Between Them ... 32

Reticulation, the Bête Noire of Cladistics ... 33

Coda ... 34

It is easy enough to collect scientific data on *Escherichia coli* 0157:H7 or *Sulfolobus solfataricus* strain P2 and to pronounce them "prokaryotes." But it is not so easy to say what the word "prokaryote" refers to, in any general sense. The issue is ontological. Does "prokaryote"' refer to a category of things existing independently of our categorizing, perhaps a "natural kind" to which any organism can be surely said to belong or not belong? And if "prokaryote" *is* such a natural kind, is it defined by its essential properties or by its genealogical relatedness to other such kinds (O'Hara 1998)? Or does the word instead denote merely a human concept, initially constructed on incomplete knowledge and possibly now hopelessly out-of-date? When we ask whether an organism belongs to some recognized and named taxon, such as Archaea, Aves, or *E. coli*, we at least have in mind both a system of classification and an understanding of a coherent theory (evolution) that gives us faith in the legitimacy of that system. But is the same true for "prokaryotes"? For "Prokaryota"?

These may seem to be abstract philosophical questions, of little relevance to practicing scientists. And yet microbiologists debate fiercely and tediously not only over what "prokaryote" means but whether or not the word should be used at all. In just the last 2 or 3 years, the dispute has taken on new heat, such that some leading scientists would surely question whether there is any reason other than misguided adherence to tradition for a volume like this, dealing with Bacteria and Archaea as if they were one kind of thing. Not least, the concerns are pedagogical. Norman Pace, an unquestioned leader in microbial evolution and systematics, has written:

> Because it has long been used by all texts of biology, it is hard to stop using the word, prokaryote. But the next time you are inclined to do so, think what you teach your students: a wrong idea. (Pace 2006, p. 289)

This chapter, then, hopes to explain why such questions remain alive, and—to some of us at least—still interesting. Our goal is exploration and explication, not resolution, because we believe that the question "What is a prokaryote?" *has* no simple answer. What Evelyn Fox-Keller has recently written about the nature versus nurture debate applies with equal force to the two domain versus three domain contretemps (and perhaps in most areas of protracted biological contestation).

> There is no single answer to this question, for a number of different questions take refuge under its umbrella. Some of the questions express legitimate and meaningful concerns that can be addressed scientifically, others may be legitimate and meaningful, but perhaps not answerable, and still others simply make no sense ... all these different questions are tangled together into an insoluble knot, making it all but impossible for us to stay clearly focused on a single, well-defined, and meaningful question. (Fox-Keller 2010, p. 1)

Some Tangled-Together Questions About "Prokaryote"

> The lessons of the three-domain tree are profound. Instead of two kinds of organism, prokaryotes and eukaryotes, there are three: bacteria, eukarya (eukaryotes) and archaea. (Pace 2006, p. 289)

At issue is whether living organisms on this planet are of two kinds (prokaryotes and eukaryotes) or instead, three (Bacteria, Archaea, and Eukarya). Begged in this pronouncement by Pace, and in fact the thorniest question taking refuge under the umbrella, is "What do we mean by *kinds of organism*"? Failure to agree on this keeps the debate alive, and no data will bring it to an end because the issue cannot be addressed scientifically, only philosophically.

E. Rosenberg et al. (eds.), *The Prokaryotes – Prokaryotic Biology and Symbiotic Associations*, DOI 10.1007/978-3-642-30194-0_114,

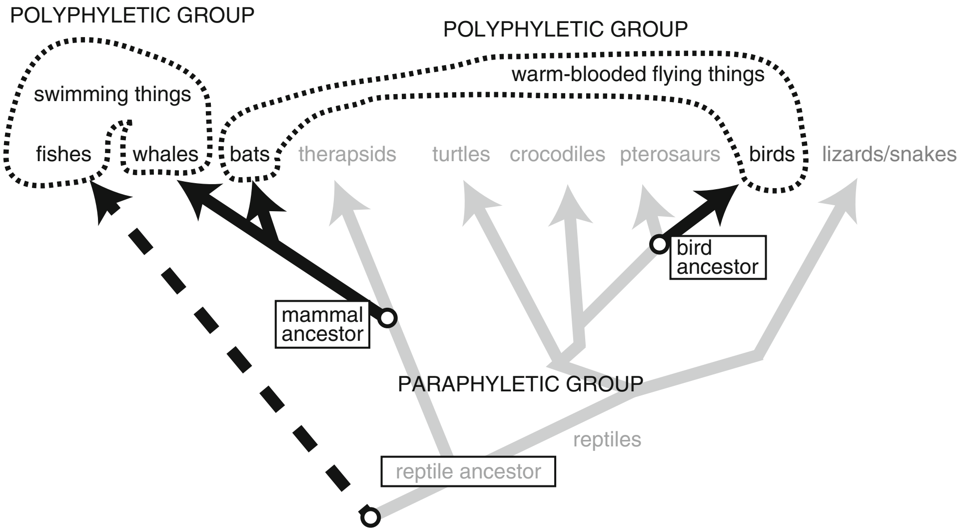

Fig. 2.1
***Polyphyly and paraphyly*. See text for explanation**

There are two general sorts of things we *might* mean by "kind" in this context, and of course an enormous relevant philosophical/biological literature, inadequately cited in the summary that follows. First, we might be thinking only in terms of current properties, as in a *phenetic* classification. We might consider prokaryotes and eukaryotes, or alternatively Bacteria, Archaea, and Eukarya, to be candidate *natural kinds*, real in a way that a category comprising all organisms taller than 1 m or whose genus name begins with the letter "E" would *not* be (Dupré 1981; Hacking 1991; Bird and Tobin 2010). They could be thought natural because they are defined by shared essences or essential properties, found in all and only representatives of the kind and somehow determinative of other characteristics. Some of the rhetoric around the two domain versus three domain controversy seems to embrace such a view, with "informational genes," especially those of translation, embodying such essential properties.

Essence-based natural kind thinking is out of fashion generally, however, and a more relaxed but still phenetic, evolution-independent, or ahistorical alternative would be that biological taxa, from species up to domains, are "cluster" kinds, which share . . .

> families of properties [that] cluster together over time either because the presence of some properties in the family favors the presence of others or because there are underlying internal mechanisms and/or extrinsic contextual mechanisms that tend to secure the co-occurrence of the properties . . . Cluster kind realists will readily concede that, depending on the case, environmental pressures may affect and alter the set of properties associated with a kind over time. Therefore, in such cases, none of the properties . . . themselves need be individually necessary for kind membership. (Bird and Tobin 2010)

Although pre-Darwinian systematists were perhaps not so wedded to an essentialistic classification as many of us were taught to believe by Ernst Mayr (Winsor 2006; McOuat 2009), it was unquestionably Darwin's *Origin of Species* that redefined the basis of naturalness in systematics, from phenetics to phylogenetics.

> All the foregoing rules and aids and difficulties in classification are explained, if I do not greatly deceive myself, on the view that the natural system is founded on descent with modification, that the characters which naturalists consider as showing true affinity between any two or more species are those which have been inherited from a common parent, and, in so far, all true classification is genealogical, that community of descent is the hidden bond, which naturalists have been unconsciously seeking, and not some unknown plan of creation, or the enunciation of general propositions, and the mere putting together and separating objects more or less alike. (Darwin 1859, p. 420)

This second, *phylogenetic* rather than phenetic way of seeing kinds as natural and real is what worked for Darwin and for most biologists today. The primary guarantor of legitimacy for any claim about how many "kinds of organism" there might be is community of descent, *not* shared current properties. We and chimpanzees are primates together because we descend from a common ancestral species that would have been considered a primate. Our many phenetic similarities with chimps justify putting us in the same order because (and *only* because) they are taken as evidence of that genealogical relationship. This second, historical or phylogenetic, way of being a natural kind might be called "tree-essentialism" and seen as a consequence of what Robert O'Hara (1997) calls "tree-thinking."

Often, classifications based on phenetic similarities and common descent or phylogeny converge. *Indeed, that they generally do was at the very heart of Darwin's case for evolution.* But they may sometimes come in conflict, in one of two ways (➋ *Fig. 2.1*). First is *polyphyly.* Insectivorous warm-blooded flying vertebrates include some birds and bats. As a taxon, "Insectivorous warm-blooded flying vertebrates" would be called *polyphyletic*, because the most recent common ancestor of birds and bats was neither warm-blooded nor flying, and possibly not insectivorous. Such polyphyletic taxa represent a serious violation of the principle of genealogical classification, and are unacceptable to biologists as "natural" taxa, however similar the organisms so aggregated might be. Whales as fishes would be an even more egregious example. If Bacteria and Archaea do *not* share a common ancestor that we could consider a prokaryote, Pace's objection to the use of "prokaryote" (as explained below) would carry considerable weight.

The second sort of phenetic/phylogenetic conflict is *paraphyly*. Birds and bats are both members of *monophyletic* taxa, in that the classes Aves and Mammalia descend from common ancestors we would consider to be class members (the last common ancestral bird and the last common ancestral mammal, respectively). Moreover, no non-birds or non-mammals descend from those respective ancestors. But the class Reptilia, from which both Aves and Mammalia arose then reveals itself to be a *paraphyletic* taxon. It includes an ancestor (the last common ancestral reptile) and some but not all of its descendants because others of those descendants are birds and others still are mammals. Cladists consider such taxa illegitimate, instead requiring that we recognize two kinds of what once were called reptiles, synapsids (including mammals) and archosaurians (including dinosaurs and birds).

But unlike polyphyletic groupings, paraphyletic taxa do not result in the false unification of species that do not share community of descent, and thus do hardly any violence to Darwin's understanding of the evolutionary foundations of natural systematics. Indeed, by recognizing that because of variations in tempo and mode some of an ancestor's descendants resemble it far less than others, the acceptance of paraphyletic taxa, coupled with knowledge of their origins, adds explanatory value to classification. Many traditional systematists, including most prominently Ernst Mayr, incorporate such considerations of what Darwin called "degree of difference" into their practice. They deploy and defend paraphyletic groupings, given adequate justification (Ashlock 1974; Grant 2003).

Birds and mammals, with respect to their reptilian ancestors, and eukaryotes with respect to prokaryotes, represent what Mayr called evolutionary *grades*.

> ▶ … the anagenetic component of evolution often leads to the development of definite "grades", or levels of evolutionary change, which must receive recognition in classification. The objection raised by the cladists that this would introduce subjectivity into classification has been rejected by the evolutionary taxonomist …. (Mayr 1982, p. 234)

As we hope to show, the argument over "prokaryote" survives in no small part because its protagonists claim to be disagreeing over scientifically addressable facts when in actuality they are arguing about different notions of what are "natural" groups, sometimes indeed from conflicting positions taken simultaneously. The grade/clade distinction is a large part of what is at stake here, but there is more.

History and Fact

> ▶ Historical narratives in which science appears to advance steadily in the direction of greater accumulations of factual knowledge are now widely scorned as "whig history." … like the stories the Whig political historians used to tell about the steady growth of English liberty. Today's historians are more likely to set themselves the goal of understanding the past "in its own terms" (whatever that might mean) rather than in the light of subsequent developments. This has yielded histories in which knowledge, rather than continuously increasing, has undergone radical discontinuities and transformations, and in which what subsequently come to be seen as forward movements are deeply rooted in contexts that are quite foreign from a modern perspective. (Golinski 1998, p. 4)

Regrettably, much history of science written by scientists is unabashedly and joyfully of the "whiggish" variety scorned by professional historians, portraying an inevitable advance (albeit with occasional detours through back alleys) from ignorance and error into truth, seldom admitting that today's knowledge may well be seen as relative ignorance tomorrow, and often overemphasizing the authors' own roles in that advance. Perhaps to working scientists accuracy and independent confirmation are essential for dealing with experimental data, but for historical "data" requiring nuanced interpretation, not so much.

Many false stories are propagated by repetitious citation, the academic equivalent of "urban myths." Worse, historical narratives can be deliberately constructed as foils against which the value of a favored new observation or hypothesis will shine out. Ernst Mayr's magisterial tome *The Growth of Biological Thought*, as we can see quite clearly now, was no disinterested exercise in pure historical scholarship (O'Malley 2010)! Moreover, even truly disinterested practitioners of a discipline will inevitably have come to whatever is their current understanding of the accumulated factual knowledge by different routes, in all honesty remembered quite differently. Outsiders cannot but be influenced most by those practitioners who write most persuasively or most often, no matter how idiosyncratic their views.

Not surprisingly, much of the dispute over the use of "prokaryote" derives from different readings of that term's intended meaning and of the common understandings of microbiologists who have used it over the last 50 years. A crucial element of the recent argument against "prokaryote" has become that it was intended – *or has come to be understood by most of its users* – to denote a phylogenetic division rather than (or as well as) a cell type, a clade rather than a grade, or a level of organization. If the former, then we would expect that many of the features in which Bacteria and Archaea are judged to be similar are genuinely homologous: if the latter, it is possible that many such features are convergent or merely analogous. Considered as a clade designation, the "prokaryote/eukaryote dichotomy" imposed upon microbiology a misleadingly reductionist, falsely unitary, and dangerously non-evolutionary worldview, its detractors claim (Pace 2006, 2009; Pace et al. 2012).

Perhaps, but a complete reading of the microbiology's past as it affects its present is surely more complex. In our view, the important and still open question about "prokaryote" concerns its value as a descriptor of a grade or level of cellular organization. That it might be also taken or mistaken as the name of a monophyletic clade (the so-called Prokaryota) is an interesting side issue. Moreover, whether or not monophyly is properly attributed to prokaryotes depends rather much on what "monophyly" itself is taken to mean, as does the assumption that any of the three domains *is* actually "monophyletic."

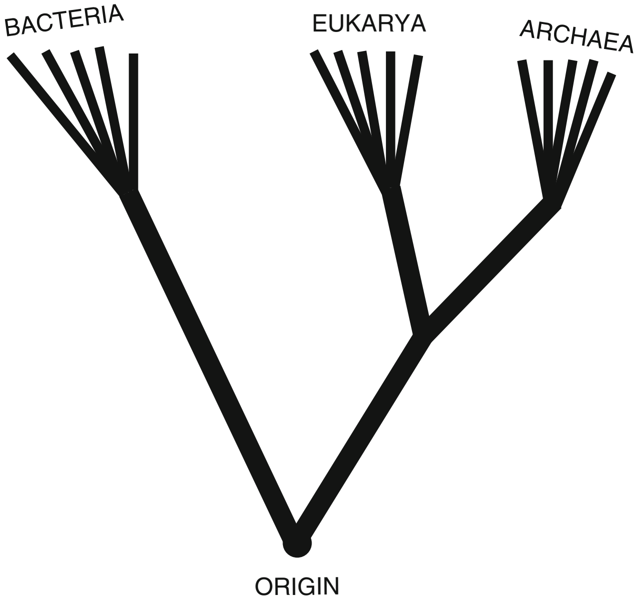

Fig. 2.2
***The three-domain tree, redrawn from* Pace (2006). Structure of the universal Tree of Life on which Pace (2006) bases his case against the use of "prokaryote"**

The Concept of a Bacterium

Jan Sapp, a professional historian unusually well-read in the microbial evolutionary/systematic literature and unusually well acquainted with the discipline's key figures, has recently highlighted some of the "urban mythology" concerning "prokaryote," while providing what seems a balanced intellectual history of the concept in "its own terms" (Sapp 2006, 2009a, b). The 1999 Berkeley Ph.D. thesis of Susan Spath (1999) earlier provided important details backgrounding the 1962 publication by Roger Stanier and Cornelius van Niel of "The Concept of a Bacterium" in *Archiv für Mikroboliogie.*

This paper is widely considered to be the cornerstone on which the belief that the most natural way to divide living things into two kinds is as "prokaryotes and eukaryotes" rested—for almost half a century. In it, Stanier and van Niel wrote:

> ▸ It is now clear that among organisms there are two different organizational patterns of cells, which Chatton (1937) called, with singular prescience, the eucaryotic and procaryotic type. *The distinctive property of bacteria and blue-green algae is the prokaryotic nature of their cells*. It is on this basis that they can be clearly segregated from all other protists (namely, other algae, protozoa, and fungi), which have eucaryotic cells. (Stanier and van Niel 1962, pp. 20–21)

Sapp shows Chatton's "singular prescience," since lauded and repetitively cited by many of us who cannot read French, to be mythic. The terms were first coined quite casually by Chatton and came to Stanier and van Niel via his most successful student, André Lwoff. But once accepted, Spath notes:

> ▸ The terminology introduced by Stanier and van Niel appears to have diffused widely through all branches of biology with little discussion . . . From the 1960s to the 1980s, the term procaryote and eucaryote were usually defined without acknowledgement of their origin. They became as much a part of standard biological discourse as 'molecule' or 'DNA.' (Spath 1999, p. 51)

Indeed, the prokaryote:eukaryote division seemed to satisfy many needs at once. Stanier and his former mentor van Niel were at pains to protect general bacteriology as a science and bacteria as a subject of study from the narrowing focus of infectious disease researchers and the reduction to mere model organisms threatened by the successes of molecular biologists who saw Bacteria primarily as tools. They hoped to draw a line above the Bacteria (separating them from eukaryotes) just as Lwoff, in his 1957 "The Concept of a Virus" had drawn a line below. Moreover, they intended to quash the false belief, supported by some microscopists, that Bacteria have nuclei. And lastly they hoped to codify a long-standing but still contested intuition that "blue-green algae"—in spite of their photobiochemical affinities to "higher" algae and green plants (and widespread speculation about their close evolutionary links thereto)— were just another kind of bacterium (indeed, "cyanobacteria").

Almost as frequently quoted as representative of Stanier and van Niel's formulation are these words from the second (1963) edition of the classic text *The Microbial World*, coauthored by Stanier with Michael Doudoroff and Edward Adelberg, which must have been written at about the same time.

> ▸ In fact, this basic divergence in cellular structure, *which separates the bacteria and blue-green algae from all other cellular organisms*, probably represents the greatest single evolutionary discontinuity to be found in the present day world. (Stanier et al. 1963, p. 85)

➋ *Figures 2.3a–d* illustrate the various ways in which these words might together be interpreted. That great "evolutionary discontinuity" might be seen (and was described in Stanier and van Niel's "The Concept of a Bacterium") as primarily a structural one, of course the product of evolution (as is all of

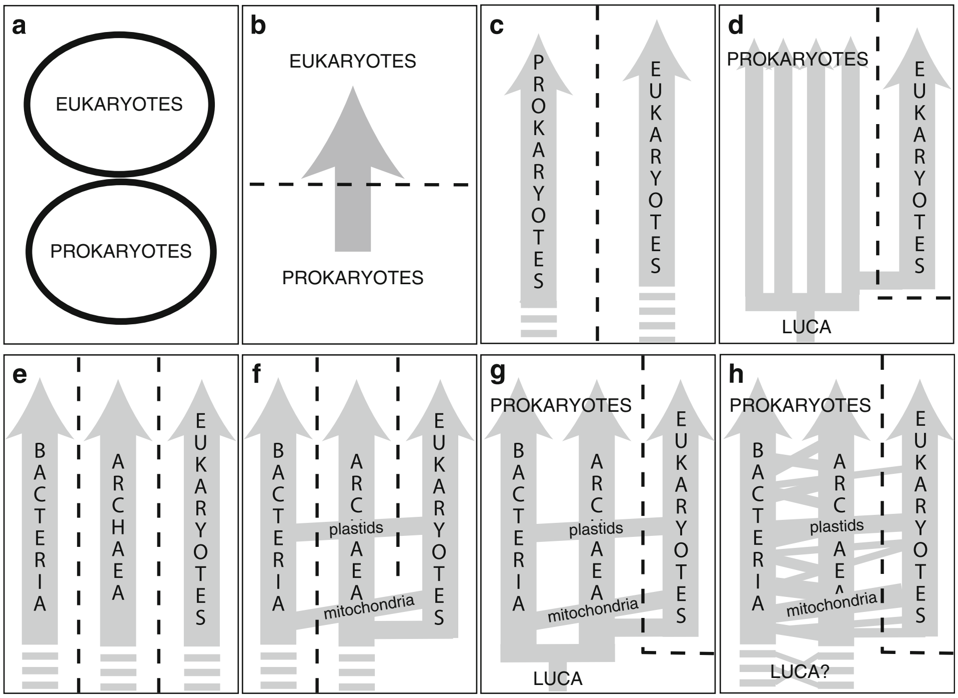

Fig. 2.3

***Eight representations of the possible relationship between prokaryotes and eukaryotes.* (a) Two distinct types of cellular organization (without implications about evolutionary relationship); (b) relationship between types, assuming that all life has a common origin and that simple precedes complex; (c) prokaryotes and eukaryotes as separately evolving clades, their common ancestor being pre-cellular, possibly as per Ernst Haeckel; (d) likely the most common view in the mid-to-late twentieth century, the Last Universal Common Ancestor (LUCA) being a cell or species of the prokaryotic type. (e) Bacteria, Archaea, and eukaryotes as separately evolving clades, their common ancestor being pre-cellular or "progenotic"; (f) as (e), but recognizing a specific Archaeal/eukaryotic ancestor, and taking into account endosymbioses giving rise to mitochondria and plastids: (g), as (f) but with the Last Universal Common Ancestor (LUCA) being a cell or species of the prokaryotic type, and Bacteria and Archaea considered a paraphyletic taxon ("Prokaryotes"); (h), as (f), but recognizing extensive interdomain lateral gene transfer (LGT). (h) seems most consistent with publications by Woese or Pace cited herein, although it is (g) that seems to be evoked in *Fig. 2.2* above**

biology), but entailing no particular phylogenetic scenario (*Fig. 2.3a*). However, if one believes that simpler cells inevitably come before more complex ones, then Stanier and van Niel's line of cell-structure demarcation had to have been crossed at least once: the prokaryotic cell type having given rise to the eukaryotic cell type (*Fig. 2.3b*). Indeed, they thought that too, and the passage quoted above continues . . .

> It is not too unreasonable to consider that the bacteria and blue-green algae represent vestiges of a stage in the evolution of cells which, once it achieved a eukaryotic structure in the ancestors of the present-day higher protists, did not undergo any further changes through the entire subsequent course of biological evolution. (Stanier et al. 1963, p. 85)

Possibly, one might consider existing prokaryotes and eukaryotes as the tips of separate evolutionary lineages, or rather clades of lineages, each with its own history and, possibly, essential characters (*Fig. 2.3c*). But if one cannot deny (as Stanier and van Niel could not) that all living things are related and believes (as they did) that prokaryotes preceded and gave rise to eukaryotes, then that "evolutionary discontinuity" also represents a deep branch in the Universal Tree of Life (*Fig. 2.3d*). The Tree's root (LUCA, or the Last Universal Common Ancestor) must have been a prokaryote.

Spath and Sapp seem to disagree on how much we are to understand "The Concept of a Bacterium" to be making a phylogenetic claim (along the lines of *Fig. 2.3d*) on top of the structural differentiation (*Fig. 2.3a*) that was so clearly its main intent. The former argues that:

> The procaryote/eucaryote distinction implicitly expressed an important taxonomic proposition about the deepest phylogenetic division among living things. At the time it was formulated, it seemed obvious to Stanier and van Niel that the procaryotes, though diverse, belonged to a monophyletic category, as did the eucaryotes. The distinction provided the basis for ending a conviction about the natural world formalized by Aristotle in the fifth century B.C. and made sacred by Linnaeus in the eighteenth century: namely, that all organisms were either plants or animals. Though challenges to that conviction had been launched since the middle of the nineteenth century, none had

> achieved wide acceptance. The procaryote/eucaryote distinction, in contrast, ended the reign of the plant and animal kingdoms as the fundamental bifurcation of living things. When compared at the cellular level, it became evident that plants and animals were much more like each other than they were like procaryotes. (Spath 1999)

Sapp (2006), on the other hand, writes that:

> ▶ Stanier and van Niel's distinction was neither an evolutionary nor a taxonomic one – at least not as they drew it. In fact, their attitude toward an evolutionary-based classification of bacteria had taken a sudden change of course prior to 1962. (Sapp 2006, p. 165)

Sapp here refers to the well-documented and often decried mid-century abandonment by these two most influential figures in microbial systematics of all hope for a satisfactorily complete phylogenetic framework for bacteriology, leaving microbial taxonomists with at best a useful "determinative," identification-directed, system of classification (see Sapp 2006; Pace et al. 2012 and the next chapter in this volume). Probably a balanced reading of "The Concept of a Bacterium" is as a sort of "at least there is this" proposition: we may never be able to sort out the tree of bacterial evolution, but we can (at least) clearly distinguish prokaryotes (of which all were then considered Bacteria) from eukaryotes, which (probably) arose from within them.

Whatever the deeper convictions of Stanier and van Niel were in 1962, many microbiologists post-1962 were indeed happy to accept the prokaryote:eukaryote split as a phylogenetic division as well as cell-structural dichotomy, as sketched in ❯ *Fig. 2.3d*. It would not however have been seen (*pace* Pace and Spath) as the deepest division in the Tree of Life, as long as one held that eukaryotes likely arose from *within* the prokaryotes, that is, from one or more already established prokaryotic lineages. So (we believe) the default notion throughout most of the last half of the last century was that it was the concomitant radical structural reorganization that made the prokaryote: eukaryote divergence into the "greatest single evolutionary discontinuity to be found in the present day living world," *not* the topology of the implicit Tree of Life (❯ *Fig. 2.3d*). For those who cared about such niceties, prokaryotes were already paraphyletic.

In any case, "The Concept of a Bacterium" taken "in its own terms" is relatively noncommittal about the specific evolutionary connection between prokaryotes and eukaryotes, while admitting to parallel evolutionary processes.

> ▶ The evolutionary diversification of the prokaryotic protists is expressed in: (1). gross organization, leading to the existence of unicellular, multicellular, and coencytic groups; (2). mode of cellular locomotion; (3). mode of cell division; and (4). major patterns of energy-yielding metabolism . . . With respect to all these features, there are parallel modes of evolutionary diversification among the eukaryotic protists (i.e. other groups of algae, protozoa and fungi). Consequently, if we look at the microbial world in its entirety, we can now see that evolutionary diversification through time has taken place on two distinct levels of cellular organization, each of which embodied, with certain limits, the same kinds of evolutionary potentialities. (Stanier and van Niel 1962, p. 33)

The Microbial World does express faith in the ability to "safely infer a common origin for the whole group [prokaryotes] in the remote evolutionary past" (Stanier et al. 1963, p. 409). Pace et al. 2012 take this as a specific inference about prokaryotic monophyly but more likely Stanier and colleagues had in mind something like ❯ *Fig. 2.3d*. They were not cladists.

One could also complain (as Pace and Woese have) that the notion that eukaryotes come from prokaryotes (❯ *Fig. 2.3b*) is progressivist and reflects an incomplete commitment to tree-thinking. Molecular biologists in particular still talk and write about a progression in complexity from prebiotic chemistry-to-bacteria-to-yeast-to-*Drosophila*-to-*Homo sapiens* in ways that call to mind *The Great Chain of Being* (Lovejoy 1936), and ignore the obvious fact that all extant lineages have equally long evolutionary histories. That prokaryotes and eukaryotes are successive rungs on the ladder of life is just another example of that pre-Darwinian way of thinking, it might be claimed. But, as argued below, if one objection to "prokaryote" is that it means nothing more than "not-eukaryote" (Pace 2006, 2009), then all that ❯ *Fig. 2.3b* or ❯ *Fig. 2.3d* imply is that eukaryotes arose from ancestors we would not now call eukaryotes. Though some authors (even one of us at one time; Darnell and Doolittle (1986)) have suggested that the first cells were eukaryote-like, with prokaryotes being their streamlined descendants (Penny and Poole 1999), this seems now an irresponsibly wild conjecture.

Three Domains Versus Two: The Debate with Ernst Mayr

As Woese masterfully summarizes in the next chapter, the proposal that living things are fundamentally of three kinds, not two, did not have the same easy ride to acceptance as did Stanier and van Niel's dichotomy—in no small part because of the widespread, almost culturally ingrained acceptance of the latter. Many of us, biologists, felt or still feel deeply that prokaryotes versus eukaryotes is "just the way it is," not from any direct observational experience or dispassionate consideration of the "accumulation of factual knowledge," but because it seems foundational to the biology we were taught and the institutional and economic structuring of our disciplines, even our social lives!

Somewhat belatedly, Ernst Mayr (1990, 1991, 1998) vigorously expounded a traditional(ist) objection to Woese's three domain view, based both on his (Mayr's) different perspective on how classification *should* be done and what he thought the Archaea were *really* like. Mayr contrasts his favored mode of taxonomic practice, "Darwinian classification" (aka "evolutionary taxonomy"), to the "Hennigian cladification" to which he thinks Woese largely, if incompletely, adheres. With the former . . .

> ▸ organisms are grouped into taxa on the basis of two criteria, similarity and genealogy. A higher taxon recognized by these criteria is composed of a group of similar and/or related species descended from their nearest common ancestor. Such a taxon is called monophyletic. In a cladification, favored by cladists, only genealogy is considered. It recognizes branches (clades) of the phylogenetic tree, comprised of the stem species of such a branch together with *all* its descendants. (Mayr 1998, p. 9721, Emphasis ours)

The important difference indeed is in the word "all." Citing the already familiar example of reptiles, birds, and mammals Mayr (1998) continues . . .

> ▸ In both cases, the cladist removes the branches that gave rise to the mammals or birds from the reptiles, thereby making the reptiles, a taxon used in our every-day grouping of animals, a "paraphyletic group," not permissible as a formal taxon in a strictly cladistic arrangement. In both cases, the Darwinian taxonomist, who deals with groups rather than with branches, retains the ancestral groups within the Reptilia and recognizes as mammals or birds only those assemblages of species which by their diagnosis are characterized as mammals or birds. It was on this basis that Stanier and van Niel recognized two empires, the prokaryotes and the eukaryotes. (Mayr 1998, p. 9721)

Mayr, like Woese and most biologists today, accepted the phylogeny displayed in ➋ *Fig. 2.3f* or ➋ *Fig. 2.3g*. But he felt that the similarities between Bacteria and Archaea were so great and their collective similarity to eukaryotes so small that the former two are legitimately classified into one superkingdom (Prokaryota, as shown in ➋ *Fig. 2.3g*) and the last into another (Eukaryota). It was a matter of recognizing that—as a grade—the eukaryote lineage had undergone massive, presumably selected, changes in the organization of its informational and especially cellular machinery. Moreover, he pointed out that Woese was not a very thorough-going cladist, because by those principles (and with the branching pattern shown in ➋ *Fig. 2.3g*) Woese *should* recognize only two primary domains, Bacteria and a second (called Neomura by Cavalier-Smith (2002)) comprising two sub-domains, Archaea and Eukarya. Mayr charitably attributed these failings to Woese's naïveté, as he expressed in a letter to WFD, dated January 31, 2000, that . . .

> ▸ Woese came into microbiology from outside of biology and did not (and still does not) understand what classification is all about. Biologists trained in classification and evolution would like to express the fact [of] how different the Prokaryotes are from the Eukaryotes.

But, one could easily rebut, Mayr did not understand what Archaea were all about! Woese, in the next chapter, emphasizes the many ways in which Archaea and Bacteria differ in basic cell and molecular biology. And in his immediate response to Mayr's 1998 attack, he cogently noted . . .

> ▸ Diversity can be of many types. It can be at the level of structure and organization; it can be anabolic or catabolic enzymatic diversity; it can be environmental adaptation at the molecular/biochemical level; it can be in the basic information processing systems of the cell; and so on. Clearly the vast diversity among birds and among insects is structural diversity, whereas that among the Bacteria or the Archaea is necessarily of the other types. Dr. Mayr's is an eye-of-the-beholder type of diversity. It rests on the incredible capacity of the human eye to distinguish minute differences in pattern. But almost all microbial diversity cannot be sensed visually, which means that subtle variations in pattern almost always go undetected... When he compares plant and animal diversity to microbial diversity, Dr. Mayr is comparing apples and oranges, and his attempt to apply globally a parochial and subjectively defined concept of diversity serves only to reveal the futility in such an approach. (Woese 1998a, p. 11045)

Time for a Change?

In 2006, Norman Pace took up the attack on "prokaryote" with renewed vigor. In an pithy one page editorial entitled "Time for a Change" appearing in *Nature* (Pace 2006), he drew attention to what he considered the implications for systematics of the rooted version of the three-domain rRNA tree (Woese et al. 1990), in which Archaea and Eukarya are sister taxa, splitting later than Bacteria. His argument as presented then and reiterated 3 years later (Pace 2009) was fourfold.

1. How, he asks, can we consider Bacteria and Archaea to form a single group (the Prokaryota) when in fact one of its two subgroups, the Archaea, is more closely related to the eukaryotes? (See ➋ *Fig. 2.2*, reproduced from Pace 2006). This would be to recognize a paraphyletic taxon (like Reptilia), anathema to cladists.
2. The pro- and eu- prefixes themselves imply that prokaryotes gave rise to eukaryotes. But "the nuclear line of descent is as ancient as the archaeal line and not derived from either archaea or bacteria."
3. Prokaryotes are defined only negatively, by those eukaryotic features that they lack. "No one can define what is a prokaryote, only what it is not."
4. "Lumping bacteria and archaea conceptually discounts fundamental differences between these two kinds of organism . . ."

Thus, he concluded that "prokaryote" embodies a scientifically disprovable false idea . . .

> ▸ Prokaryote: gene-sequence comparisons show the tree of life consists of Bacteria, Eukarya, and Archaea. The use of the term "prokaryote" fails to recognize that an idea about life's origins has been proved wrong. (Pace 2006)

There is much that can be said in rebuttal and was, for instance, in correspondences to Nature by Martin and Koonin (2006), Dolan and Margulis (2007), and Cavalier-Smith (2007) and later by Whitman (2009) in the *Journal of Bacteriology.* An excellent recounting of the published and unpublished contestation following "Time for a Change" is provided by Sapp (2009a, b). Each of Pace's four complaints is debatable at length, and is the subject of one of the next four sections of this essay.

Paraphyly

Pace's first argument seems to be that *prokaryotes comprise a paraphyletic group*. This should thoroughly distress only cladists, which neither Pace nor Woese in other contexts appears to be. Their above-mentioned refusal to recognize or name Neomura as a clade shows that they are not strict Hennigians, as does their insistence on the meaningfulness of "domain" status. Degree of difference (especially in the informational transcription, translation, and replication machineries) and the fact that Archaea and Bacteria are separated by deep and long branches in the rRNA tree both figure in Pace and Woese's case for the domain status of either, and are most consistent with the practices Mayr styles "Darwinian classification."

Most microbial systematists, and indeed most biologists who care at all about microbes would be aware of the benefits and dangers of the grade/clade distinction. But many of them would argue that the elaboration of the cytoskeleton and endomembrane systems and the uncoupling of transcription and translation that seemed to have occurred rapidly and simultaneously in the formation of LECA (the last universal eukaryote ancestor), however that came about, were so radical and of such consequence as to make the prokaryote-eukaryote divide still "the greatest single evolutionary discontinuity to be found in the present day world" (Embley and Martin 2006; Field and Dacks 2009). Thus, as Whitman (2009) writes,

> ▸ Even though the prokaryotes are not monophyletic and the evolutionary processes they have experienced are extremely complex, this classification strategy remains useful and knowledge of the evolutionary processes which formed modern organisms provides a great deal of insight into their biological properties. (Whitman 2009, p. 2003)

"Pro" Implies Before

Pace's second objection, *that the "pro" in prokaryotes is misleading because it implies an ancestral relationship to eukaryotes*, seems inconsistent with his first, almost disingenuously so. To call a taxon paraphyletic is to admit that another taxon emerged from within it. It is of course commonly understood that contemporary taxa did not evolve from each other. Humans did not evolve from chimpanzees: they instead share with chimps a common hominid ancestor, which was neither chimp nor human. But one could be almost certain it was hairier than us, because all the outgroup apes are. Similarly, the common ancestor of Archaea and eukaryotes was by definition neither one nor the other. But simple parsimony together with the accepted three-domain tree dictates that features found in eukaryotes, but not in Archaea and Bacteria, are more likely to have been invented in the eukaryotic branch than lost in the archaeal branch. And if, as Pace observes, all it takes to be a prokaryote is to lack eukaryotic defining traits, then the common archaeal/eukaryote ancestor surely *was* a prokaryote, although (if Archaea and Eukarya are indeed sisters) more Archaea-like than Bacteria-like.

Moreover, it seems now equally possible that the eukaryotic "nuclear line of descent" is not in fact "as ancient as the archaeal line," but arose from within it! Such an alternative view has been around almost as long as Archaea have been recognized, James Lake having suggested in the early 1980s, on the basis of ribosome structure, that eukaryotes show a particular affinity to one of the two then recognized archaeal subdomains, the Crenarchaeota, which he called "eocytes" (Lake et al. 1984) Martin Embley and coworkers (Cox et al. 2008; Foster et al. 2009) find that sophisticated and conservative phylogenetic methods that compensate for compositional and rate heterogeneity across sites and across trees are indeed prone to produce the "eocyte tree" in preference to the accepted three-domain tree, when applied to rRNA gene sequences or to a set of several dozen highly conserved and "core" genes of translation, transcription, and replication.

Taking into account also the presence or absence of certain key genes—several ribosomal protein genes, two RNA polymerase subunit genes, genes for components of the cell division machinery and a ubiquitin modifying system, and most excitingly a cell-shape-determining actin ortholog—Guy and Ettema (2011) recently argued for a specific placement of eukaryotes within what they call the TACK superphylum of Archaea. This comprises Crenarchaeota, Thaumarchaeota, Korarchaeota, and the recently proposed Aigarchaeota, and specifically excludes Euryarchaeota and Nanoarchaeota. Given the recent and still unstable taxonomic treatment of an expanding archaeal data set, we take this as being a modern version of Lake's "eocyte hypothesis." If this view gains wide acceptance and Archaea themselves become paraphyletic, Pace's phylogenetic/cladistic argument against prokaryotes would also deconstruct Archaea. At the very least we should remain agnostic as to the relationship between Archaea and the eukaryotic nuclear lineage (Gribaldo et al. 2010).

Negative Definition

Pace's third argument, *that prokaryotes are defined only negatively*, has more bite, and requires more discussion. Stanier and van Niel's definition of "prokaryote" was indeed dismayingly minimalistic.

> ▸ The principal distinguishing features of the procaryotic cell are: 1. absence of internal membranes which separate the resting nucleus from the cytoplasm, and isolate the enzymatic machinery of photosynthesis and of respiration in specific organelles; 2. nuclear division by fission, not by mitosis, a character possibly

> related to the presence of a single structure which carries all the genetic information of the cell; and 3. the presence of a cell wall which contains a specific mucopeptide as its strengthening element. (Stanier and van Niel 1962, pp. 32–33)

Even these "negative" descriptors now seem problematic, since none is thought true of *all* prokaryotes. The specific peptidoglycan mucopeptide referred to is absent from eukaryotes (except in some photosynthetic organelles of cyanobacterial origin), but also from many bacterial and all archaeal walls, and there is much compositional variation in the walls of both prokaryotic domains (Whitman 2009). Fission is not the only way either Bacteria or Archaea divide (Angert 2005; Makarova et al. 2010). And although no prokaryote has internal membranes that are likely to be homologous to eukaryotic membranes, it is the case that within the planctomycetes there are remarkable *analogs* thereof (Fuerst and Sagulenko 2011), arising by covergence or gene transfer (McInerney et al. 2011).

Still, it is not always so easy to tell negative from positive in such general descriptions of organismal characteristics. The absence of functionally differentiated internal membranes has, for instance, been effectively re-cast positively, as the multifunctionality of the prokaryotic cytoplasmic membrane, by Whitman (2009).

> The cytoplasmic membrane is multifunctional in procaryotes and represents the defining structure of the cell. A proton motive force is generated on the cytoplasmic membrane by respiration, photosynthesis, or ATP hydrolysis to empower key cellular processes such as ATP biosynthesis, NAD^+ reduction by reverse electron transport, nutrient uptake, motility, and secretion. Procaryotes utilize membrane transporters on the cell surface to assimilate nutrients dissolved in their environment. In many procaryotes, the cytoplasmic membrane possesses a complex topology composed of lamellae, tubules, or other cytoplasmic intrusions. In contrast, the cytoplasmic membrane of eucaryotes is very different in structure and function. (Whitman 2009, p. 2000)

Similarly, Martin and Koonin (2006), in an immediate response to Pace's challenge in *Nature*, point out that the absence of nuclear membranes makes possible the coupling of transcription and translation and several prokaryote-specific types of regulation, a seemingly "positive" feature. Coupling might necessitate the absence of spliceosomal introns, which could be "positivized" for prokaryotes as "having continuous genes." Coupling might also mandate more direct forms of regulating gene expression (Payankaulam et al. 2010). Indeed, there is considerable literature indicating a common prokaryotic repertoire of transcriptional regulatory factors (Bell and Jackson 2001; Peeters and Charlier 2010).

There is also an enormous shared prokaryote-specific pool of "operational" genes, as illustrated in ➲ *Fig. 2.4* and previously in Walsh and Doolittle (2005). Thus, though (1) Bacteria and Archaea are indeed distinguishable from eukaryotes and each other by the character of the "informational genes" of transcription, translation, and replication, and (2) for many of *these* genes a specific archaeal/eukaryote relationship may be demonstrated (Cox et al. 2008; Cotton and McInerney 2010), it is still true (3) that a much larger number, many hundreds, of "operational" genes make up a shared resource that is common to Bacteria and Archaea. Such genes are often (on an evolutionary time scale) laterally transferred within and between them, and generally absent from eukaryotes. Indeed, one might define "prokaryotes" "positively" on this basis. They are that community of organisms that, over an evolutionary timescale, draws on a vast shared pool of genes encoding diverse metabolic functions seldom if ever used by eukaryotes. In a sense, evolution by LGT remains an important defining feature of prokaryotes, just as Woese would have it have been for progenotes, a feature which eukaryotes lack, relatively speaking (Woese 1998b, see below).

In any case, the key systematics question is not so much whether there are "non-negative" features uniting Bacteria and Archaea as whether there is *any* reason to believe that they derive from a common ancestor that might itself be called a prokaryote, so that they might together be considered a legitimate clade—albeit a paraphyletic one. There are two different questions embedded here: (1) what do we mean by common *ancestor* (versus common *ancestry*), and (2) do we have a sufficiently refined definition of "prokaryote" to be able to say whether or not the common ancestor, if it existed, was one? As with nature versus nurture, "these different questions are tangled together into an insoluble knot" (Fox-Keller 2010).

In several earlier and recent writings, Woese has argued that the common ancestor of all three domains was not a prokaryote or even a population of prokaryotes, but rather "the progenote," an inchoate collection of entities still in the throes of evolving the genotype/phenotype coupling. The manifest differences in information processing machinery between Bacteria, Archaea, and eukaryotes tell us about that early period, he feels, and suggests a model in which the three domains are best thought of as arising independently from an ancestral "state" rather than an ancestral cell or species, as in ➲ *Fig. 2.3e*. Of that state, he wrote (Woese 1998b)...

> Organismal lineages, and so organisms as we know them, did not exist at these early stages. The universal phylogenetic tree, therefore, is not an organismal tree at its base but gradually becomes one as its peripheral branchings emerge. The universal ancestor is not a discrete entity. It is, rather, a diverse community of cells that survives and evolves as a biological unit. This communal ancestor has a physical history but not a genealogical one. Over time, this ancestor refined into a smaller number of increasingly complex cell types with the ancestors of the three primary groupings of organisms arising as a result. (Woese 1998b, p. 6854)

It surely is the case that evolution *was* different in tempo and mode in the ancient pre-cellular past. One of us (WFD; Doolittle and Brown 1994), called this the period of "progressive Darwinian evolution" because many improvements in the speed, accuracy, and efficiency of replication, transcription, and translation that would have tightened the genotype-phenotype coupling and helped cells adapt to environmental

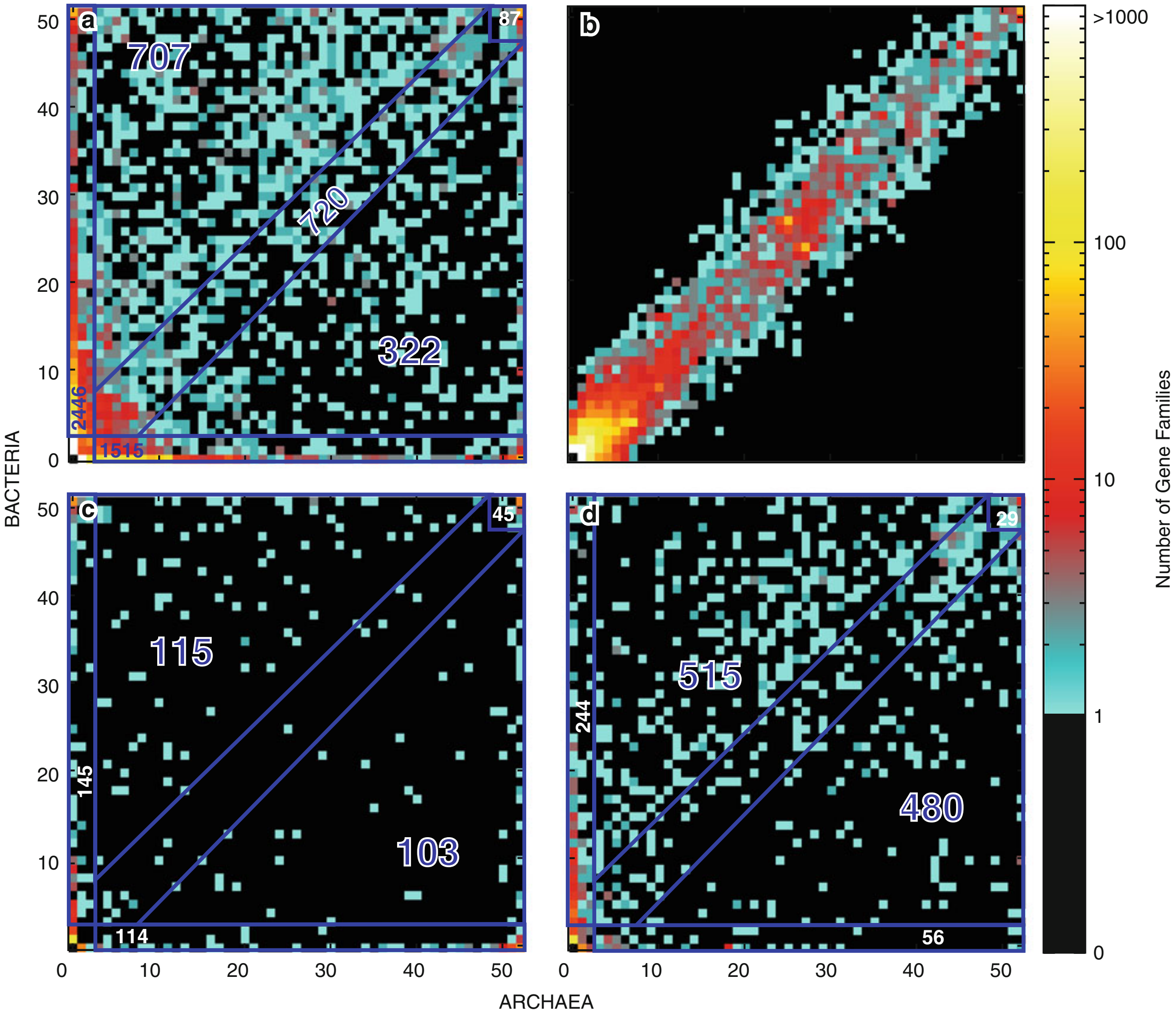

Fig. 2.4

Distribution of gene families in representative archaeal and bacterial genomes. To reconstruct gene families, we performed all-against-all BLASTP (E-value cutoff of 10^{-4}) searches (Altschul et al. 1997) of open reading frames (ORFs) in 103 selected genomes. Normalized bitscores from pairwise BLASTP comparisons were used to cluster ORFs into superfamilies using MCL algorithm (inflation parameter I = 1.1; Enright et al. 2002). Superfamilies with more than four members were further divided into gene families using the BRANCHCLUST pipeline (Poptsova and Gogarten 2007). To select representative genomes among completed sequenced genomes available in GenBank we used the following criteria: (a) Only genomes larger than 1 Megabases were considered; (b) The largest representative per Archaeal genus was taken, resulting in 52 genomes; (c) The largest two representatives per bacterial phylum representing at least different genera were chosen. For broadly represented phyla (such as proteobacteria and cyanobacteria, among others), additional genomes were added until the number of selected bacterial genomes was comparable to the number of Archaeal ones (total of 51 genomes). The obtained gene families' distribution is visualized in a heat map, where the number of gene families present in *x* archaeal and *y* bacterial genomes, shown on X and Y axes, respectively, is color coded on a logarithmic scale (see color legend on the *right* of the figure; *black* values corresponds to 0 families, and *white* to >1,000 families). *Numbers* depicted over a heat map correspond to the total number of gene families in regions, perimeters of which are delineated by *blue lines*. *Panel A*. Comparison of all gene families. Near-universal core consists of 87 gene families. While the bulk of gene families present in at least four genomes can be called domain-specific (2,446 for Bacteria and 1,515 for Archaea), only very small fraction of those is universally present in the majority of the genomes of each domain. A substantial number of genes (1749) are patchily distributed (shared) across the two domains. *Panel B*. Randomization of 103 genomes (one of five replicates is shown in this panel) into two groups resulted in a different distribution of genes, suggesting that the distribution of gene families among the archaeal and bacterial domains is not random. *Panel C*. Subset of ''informational'' gene families (as defined by ''Information Storage and Processing'' functional category of the Clusters of Orthologous Groups, or COG, database (Tatusov et al. 2003)). *Panel D*. Subset of ''operational'' gene families (as defined by ''Metabolism'' functional category of the COG database) challenges of all sorts should have been under positive selection then. True "progess" was being made: organisms would have been getting better just at being organisms. After that (after crossing what Woese calls "the Darwinian threshold"), adaptations would have become more ecology-driven, tracking environmental change—even "reversing" themselves, and progressive only in a highly contingent and local context.

But whenever and at whatever stage we consider the common *ancestor* of Bacteria and Archaea to have appeared, or indeed even if we think there was no "discrete entity" deserving of that name, there is no denying that the two domains share a common *ancestry.* The universality of the genetic code of the ribosome, and of many of its constituent proteins (Fox 2010), of demonstrably homologous RNA polymerase subunits (Werner and Grohmann 2011), and of so many operational enzymes and the biochemistries they promote make it impossible to imagine an independent origin of Bacteria and Archaea. That is, they could not have gone from prebiotic chemistry through to the first cellular forms that we might consider Bacteria or Archaea without extensive pooling of information. Even Bill Martin and Mike Russell (2003)—whose origin-of-life scenario entails advancement to the stage of ribosomes and the universal code among *pre-cellular* systems maintained in FeS compartments—must invoke extensive information exchange.

Vetsigian et al. (2006) describe the code as the product of "collective evolution" via LGT…

- The phylogenetic expression of ambiguity is reticulate evolution. In reticulate evolution, there is no unique notion of genealogical descent: genetic content can be distributed collectively… The players are cell-like entities still in early stages of their evolutions. The evolutionary dynamic (the "rules") involves communal descent. The key element in this dynamic is innovation-sharing, an evolutionary protocol whereby descent with variation from one "generation" to the next is not genealogically traceable but is a descent of a cellular community as a whole. Even if an organismal ancestry were to some extent traceable, it would have no significance, because it is the community as a unit, not the individual organismal lineages therein, that varies in descent… The central conjecture in our model is that innovation-sharing, which involves horizontal transfer of genes and perhaps other complex elements among the evolving entities [a dynamic far more rampant and pervasive than our current perception of horizontal gene transfer (HGT)], is required to bring the evolving translation apparatus, its code, and by implication the cell itself to their current condition. (Vetsigian et al. 2006, p. 10696)

Importantly, and perhaps ironically, this "dynamic" is also how one might describe the evolution of a complex trait within a modern multicellular sexually reproducing species, ours for instance. Because of recombination, to which mammalian reproduction is inextricably tied, any multigenic advance is similarly going to be the product of "innovation sharing." Though each contributing mutant allele should be traceable to a last common ancestor (its "coalescent") arising earlier in the population, the complex trait as a whole will not be "genealogically traceable," and is indeed the achievement of the "community as a whole." The difference of course is that with *Homo sapiens* it is homologous recombination *within* the species that underwrites the process, whereas for prokaryotes or the "cell-like entities" comprising the progenote, it is LGT (alternatively HGT) *between* "species."

Both within-species recombination and between-species LGT/HGT complicate the concept of "ancestor" (as opposed to ancestry), in a way not often appreciated by microbial phylogeneticists, who tend to equate gene trees with species trees. In a recombining sexually reproducing species, alleles at different loci find their last common ancestors (their coalescents) in different genomes present at different times in the population's past. "Gene trees" are uncoupled from "species trees" and sexually reproducing organisms have organismal ancestors from whom they have received no genetic information. (We receive half our alleles from each parent but it is possible if ridiculously improbable that we received none from one of our grandparents, and in fact certain that, many more generations back, we have reproductive ancestors from whom we received no alleles.)

Similarly, because of LGT, the last common ancestral versions of the many gene families found distributed in Bacteria and Archaea today will trace to common ancestral versions that existed in different genomes at different times in the history of their ancestral global prokaryotic population. The last common ancestral prokaryotic cell quite possibly housed none of them. Some researchers think that a small core (as few a 1%) of all genes, mostly involved in translation, have never been subjected to LGT, and thus do find their last common ancestral versions in the genome of a single cellular Last Universal Common Ancestor, or LUCA (Ciccarreli et al. 2007). Others, who may doubt the existence of any such never-transferred genes but still hope to give a traditional meaning to the Tree of Life, assert that, rather than being a single individual or even a single species, "LUCA was a population" (or ancestral "state"), indeed a heterogeneous population extended through time, possibly over the many tens or hundreds of millions of years consumed by "progenotic," or "progressive Darwinian" evolution (Koonin 2009; Glansdorff et al. 2008, 2009; Di Giulio 2011).

This seems a radically different use of the word and concept of "ancestor," which is more commonly understood to single out *individuals* in earlier populations. And we need not buy into such a redefinition (or indeed any definition of common *ancestor*) in order to accept a general model of common *ancestry,* in which the various features of modern cells evolved "collectively" (i.e., in no single genomic lineage), in a globally communicating (via LGT) super-population of entities becoming, over the course of hundreds of millions of years, ever more like contemporary prokaryotic cells.

A recasting of the question asked earlier, whether there is *any* reason to believe that Bacteria and Archaea derive from a common ancestor that might itself be called a prokaryote, is this. "Would that ancestral cell or cells in which the last common ancestral versions of genes that (it is assumed) have never been laterally transferred – in particular ribosomal RNA and protein genes – be more “primitive” in terms of its informational machineries (their accuracy, efficiency or general character) than are contemporary prokaryotes?” This is a formulation that we think most authors who embrace either a single-cell or heterogenous population notion of LUCA might accept. We submit that we do not, in spite of the great appeal of the progenote concept and the various ingenious but tortured arguments put forth in favor a primitive, even RNA-genomed ancestral state (Koonin 2009; Glansdorff et al. 2008, 2009; Di Giulio 2011), know the answer to it.

To be sure, Woese’s observation that the many structural differences between the informational machineries of Bacteria and Archaea, particularly the ribosomes, reflect their independent evolutionary refinement from a more primitive (less accurate and efficient) ancestral form is the best-articulated argument along these lines. In concluding a recent survey of domain-specific features (“signatures”) of bacterial and archaeal ribosomes, he and colleagues (Roberts et al. 2008) wrote . . .

> ► Through our analysis of ribosomal signatures, we have provided a glimpse into the evolutionary past, at the “base” of the [universal phylogenetic tree]. This study has identified the ribosomal signatures and provided examples of how they are helpful in understanding the evolutionary dynamic by which the ribosome arose. These signatures give each phylogenetic domain a distinctive character and bespeak stages through which the evolution of the ribosome must have proceeded, both before the emergence of the individual lineages themselves (in the universal ancestral state) and subsequently, separately within each primary lineage. (Roberts et al. 2008, p. 13958)

But the differences between archaeal and bacterial ribosomes and the very long branches between bacterial and archaeal domains in rRNA and many other gene trees are not *necessarily* evidence of a different mode and tempo of evolution or of a primitive LUCA at the root of any gene tree. All bifurcating trees must have a deepest branching that divides them into two subtrees. On average (assuming a random birth-death model) the two branches will themselves appear unbranched halfway up the tree, this being the null model under coalescence theory (Zhaxybayeva and Gogarten 2004). Unusually rapid evolutionary rates and ancestral primitivity need not be invoked.

Moreover, there are selective (Gupta 2000) and nonselective (Lukes et al. 2011) forces other than improvement from a primitive state that might radically transform the translational machinery. Mitochondrial ribosomes, for instance, are vastly different in composition and structure from their alpha-proteobacterial ancestors (O’Brien 2002), and eukaryote ribosomes are larger and more complex than their presumed archaea-like predecessors. If we accept the scenario in ➋ *Fig. 2.3e*, these latter differences might bespeak separately evolved refinements of a primitive progenotic ancestral ribosome. But most would favor phylogenies more like those in ➋ *Fig. 2.3f–h*. We can infer from the similarity across bacterial phyla that the bacterial ribosome was already in the final form at the time LECA (the last eukaryote common ancestor) welcomed on board the alpha-proteobacterial symbiont that was to become the mitochondrion. And evidence favoring the “eocyte” hypothesis (Cox et al. 2008; Foster et al. 2009) also implies that LECA’s ribosomes were already fully functional archaeal types.

We cannot from comparative ribosomology alone decide whether the ancestral cell or cells in which the last common ancestral versions of ribosomal RNA and ribosomal protein genes resided were “progenotes” or “prokaryotes.” If progenotes, then “prokaryote” becomes polyphyletic, as well as paraphyletic, and the word would have two strikes against it. But this second strike depends entirely on how we define both “prokaryote” and “progenote,” not “meaningful concerns that can be addressed scientifically” and on the existence and nature of LUCA, which is similarly problematic.

Bacteria and Archaea, and the Fundamental Differences Between Them

Pace’s fourth argument is that *disregard of these differences “reinforces an incorrect understanding of biological organization and evolution”* (Pace 2006). To be sure (rephrasing Stanier et al. 1963) the “basic divergence in cellular structure which separates the bacteria [and archaea] from each other represents the greatest single evolutionary discontinuity to be found in the present day [prokaryotic world].” In the next chapter, Woese summarizes those differences—in membrane composition and the machineries of translation, transcription, and replication—that have long been considered diagnostic of the bacterial/archaeal discontinuity. It is also the case that “operational” genes (primarily those of catabolism and anabolism), though widely shared between Bacteria and Archaea, are nevertheless often preferentially associated with one or the other domain. The gene family presence/absence data presented in ➋ *Fig. 2.4* and the comparison to a random partitioning of genes make this abundantly clear.

The “incorrect understanding” that concerns Pace is the presumption of uniformity among prokaryotes. Stanier and van Niel were themselves cognizant and celebratory of bacterial diversity (Spath 1999; Sapp 2006, 2009a, b), though they were not aware of the “discontinuity” between Bacteria and Archaea (nor was anyone at the time). It seems unlikely that they would have been dismissive of the kinds of evidence Woese brings to bear. Indeed, Woese lays much of the blame for what he sees as a woeful neglect of difference on molecular biologists, whose reductionist paradigm led them to think that what is true for *E. coli* is true for elephants and (coincidentally) for *Bacillus subtilis* and *Sulfolobus solfataricus*.

It is not just the bacterial/archaeal distinction that is painted over by the broad brush of molecular biology, however. There are fungi whose molecular genetics are wildly unlike that of *Saccharomyces cerevisiae*, insects whose population genetics are not like that of *Drosphila melanogaster*, vertebrates whose development is strikingly different from that of *Gallus domesticus*, and primates who do not behave much like us. It is not surprising that biologists who have spent their lives with one model organism want their results to be seen as more generally valid and that cell and molecular and cell biology textbooks follow their lead, while teachers of evolution and diversity defend differences but reach fewer students.

As Woese himself noted in rebuttal to Mayr, "Diversity can be of many types," and what we emphasize will depend on our experience, disciplinary focus, and research goals. That Bacteria and Archaea appear different in many fundamental biological features is beyond question: whether this degree of difference is dwarfed by that between prokaryotes and eukaryotes is not something that science can decide. This is, as Woese noted in the passage quoted earlier, to compare apples to oranges.

Reticulation, the Bête Noire of Cladistics

Trees based on rRNA are unrootable, because there is no outgroup. The accepted rooted version of the three-domain rRNA tree (as in ➋ *Fig. 2.2*) was initially based on a few widely distributed paralogous protein-coding gene families, assumed to be the products of gene duplications that predated LUCA (Iwabe et al. 1989; Gogarten et al. 1989; Brown and Doolittle 1994). The tree based on sequences of one of the duplicates can be used to root that of the other and vice versa, the two trees being in principle congruent with each other and with the true Tree of Life (assuming no LGT). Such duplication-based rootings generally showed eukaryotes and Archaea to be sisters, an inference supported by strong and specific similarities between elements of the archaeal and eukaryotic replication, transcription, and translation machineries (Pace et al. 2012).

Of course it was known since the early seventies that there would be *some* bacterial contribution to the eukaryotic nuclear genome, via the mitochondrial symbiosis and organelle-to-nucleus gene transfer. The initial assumption that this contribution would be small—and perhaps a prejudice that it is the phylogeny of the host that matters most—led to the now largely unquestioned representation of eukaryotes as sister to Archaea.

But unexpectedly, estimates of the bacterial contribution have grown enormously as genomic data pour in. In 2004, Esser et al., described a surprising result. . .

> ▸ . . . approximately 75% of yeast genes having homologues among the present prokaryotic sample share greater amino acid sequence identity to eubacterial than to archaebacterial homologues. At high stringency comparisons, only the eubacterial component of the yeast genome is detectable. Our findings indicate that at the levels of overall amino acid sequence identity and gene content, yeast shares a sister-group relationship with eubacteria, not with archaebacteria, in contrast to the current phylogenetic paradigm based on ribosomal RNA. (Esser et al. 2004, p. 1643)

More recently, Pisani et al. (2007), using multiple (including photosynthetic) eukaryotic genomes and a larger collection of bacterial and archaeal genomes reached a similarly startling conclusion. . .

> ▸ . . . there are three distinct phylogenetic signals in eukaryotic genomes. In order of strength, these link eukaryotes with the Cyanobacteria, the Proteobacteria, and the Thermoplasmatales, an archaebacterial (euryarchaeotes) group. These signals correspond to distinct symbiotic partners involved in eukaryote evolution: plastids, mitochondria, and the elusive host lineage. (Pisani et al. 2007, p. 1752)

Why, if the Archaeal signal is the weakest, is the Archaeal/eukaryote sister relationship enshrined in textbooks, and the concepts of paraphyly employed so enthusiastically by Pace in his effort to discredit "prokaryote"? Of course, even if the preponderance of data were all that mattered, prokaryotes would still be paraphyletic, with eukaryotes seen as sisters to the Cyanobacteria or Proteobacteria. But the more we come to see eukaryotes as a chimeric clade, established by symbiosis, cell fusion, and LGT, through the active participation of several or many prokaryotic lines, the more aptly descriptive the "eu" and "pro" in "eukaryotes" and "prokaryotes" begin to look. This would have been all along the position of the late Lynn Margulis, coincidentally (Margulis 1996).

In any case, the language of cladistics and the tree-like model on which it is based seem almost irrelevant in the face of such reticulation, or non-tree like signal.

> ▸ Reticulation is thus the *bête noire* for cladistics, as initially recognized by Hennig. There are a number of sources of homoplasy (incongruency between certain character distributions and the cladogram based on maximum parsimony), such as adaptive convergence, gene conversion, developmental constraints, mistaken coding, and reticulation. The last-named factor is the most problematical because it involves the fundamental model of reality underlying cladistic analysis. (Mishler and Theriot 2000, p. 50)

Indeed, in 1975, Peter Sneath, a microbiologist and leading theoretical systematist, noted that "reticulate evolution requires consideration of ways to represent the descent of genes, instead of entire genomes as in simple branching cladograms" (Sneath 1975). In 1992, the systematist David Mindell eloquently defended

> ▸ . . . the idea that the history of no one symbiont should take precedence over another in assessing genealogy (monophyly) and classification, regardless of size, weight, status as host, or other measure of relative dominance . . . To discount one symbiont within a holobiont when reconstructing the holobiont's overall phylogenetic history and proposing a classification . . . is equivalent to disposal of data. (Mindell 1992, p. 57)

If monophyly is to retain its meaning as containing all descendants of a single ancestor and that ancestor—*and* no data are to be disposed of—then eukaryotes are not monophyletic, nor are Bacteria or Archaea, because of LGT.

A compromise that might allow for a principled if not unarguably "natural" classification in the face of reticulation was suggested in 2007 by David Baum, a noted plant systematist.

> ▸ A primary concordance tree, a tree built from clades that are true of a plurality of the genome, provides a useful summary of the dominant phylogenetic history for a group of organisms.

> Residual historical signals can also be extracted in the form of secondary, tertiary, etc. concordance trees, which are built from clades that are present in the genome but contradict clades on the primary concordance tree. (Baum 2007, p. 417)

Most of us would react with horror to a proposal to consider reclassifying non-photosynthesizing eukaryotes as sister to proteobacteria and photosynthetic eukaryotes as sister to cyanobacteria (though this last view was common until the mid-twentieth century). But this would be arguably more consistent with Darwin's call for *genealogical* classification, and would be based on the same facts and the same understanding of what has happened during the evolution of genes, genomes, and organisms as is the "phylogenetic classification" enshrined in the hegemonic three-domain tree.

It is only by "disposal of data"—allowing rRNA and a few associated informational genes to stand in for an evolutionary lineage as a whole—that inferences about monophyly, polyphyly, or paraphyly of the three domains can be made to appear clean. Two assumptions lie behind this maneuver. First is that informational genes are less likely to be transferred, especially across large phylogenetic distances, because of the complexity of the interaction of their products with those of other genes (Jain et al. 1999). Thus, they track the history of individual cell divisions and speciation events better than other genes. The second, more presumption than assumption, is that it is in this history that a true genealogical classification most naturally rests.

The first is indeed plausible from several perspectives. Interestingly, Cotton and McInerney (2010) recently published a paper whose title says much about the privileging of the archaeal trace. In "Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of function" they show that those (quantitatively) fewer yeast eukaryotic nuclear genes that show archaeal sisterhood are disproportionately "essential to yeast viability, are more highly expressed, and are significantly more highly connected and more central in the yeast protein interaction network," as might be expected if the archaeal heritage were more stably associated with the less frangible aspects of cell biology.

The second, that even when a majority of genes contradict the informational signal, it is the latter we should rely on in classification, is not a question of science, answerable by experiment. Classification has less to do with identifying "natural groups" and more to do with the philosophy, history, and politics of the idea of "natural" than we generally like to admit. There is much intellectual momentum behind this approach, but there *are* serious alternatives (Cavalier-Smith 2002, 2007) that cannot be dismissed out of hand, simply because they have fewer adherents. There is in fact no principled way, given massive reticulation, of deciding what is the most "natural" way to classify organisms, or divide all living things up into "kinds." The various positions taken by microbiology's leaders, though bolstered by appeals to scientific evidence, are in the end rhetorical.

Coda

Natural kind thinking has been part of biological classification for millennia, of course, as many philosophers of biology and biologists who practice philosophy have noted. Ernst Mayr called it "typological thinking." The most extreme form of such thinking has been described (and then rejected) by the philosopher Ian Hacking as the belief that: . . .

> ► There is a unique best taxonomy in terms of natural kinds, that represents nature as it is, and reflects the network of causal laws. We do not have nor could we have a final taxonomy of anything, but any objective classification is right or wrong according as it captures part of the structure of the one true taxonomy of the universe. (Hacking 1991, p. 111)

Such a view, which we too reject, underlies the debate about whether there are three rather than two kinds of living things in this world. To be sure, there are more or less reasonable ways to look at diversity. But as to how many kinds there *actually* are, there should be no more definitely true answer than there would be to "How many kinds of people are there in North America?" or "How many kinds of books are there in my town's public library?"

And yet, such natural kind thinking comes naturally to biologists, generating much of the heat in our arguments. Insistence that the grade distinction between prokaryotes and eukaryotes must be discarded because it conflicts with the division of living things into clades reflects a belief that "the one true taxonomy of the universe" is the Tree of Life—to which concept, as loyal Darwinists, we feel we must cling. And yet, because of chimerism and extensive LGT, that Tree means much less as a description of what organisms are like and how they came to be that way than Darwin himself would have ever imagined (Doolittle 1999; Martin 2011).

Furthermore, the Tree unambiguously depicts three discretely defined monophyletic clades if and only if we accept as settled many highly arguable propositions about how evolution proceeds, about what the data say in that regard, and about how classification is properly to be pursued. Some of these propositions, reviewed here, are that (1) in spite of the testimony of most of their genes, the "one true" way to look at Eukarya is as sisters to or emerging from within Archaea; (2) in spite of very sophisticated recent analyses of those genes thought to be most truthful, sisterhood (rather than emergence from within) is the right way to understand this relationship; (3) Bacteria and Archaea are themselves properly described as monophyletic clades, even though it is but a few percent of their genes that tell us this, with the preponderance of genes saying that they comprise a single, albeit highly structured and admittedly *very* slowly mixing population; (4) that it makes sense to speak of LUCA, either as a single cell or species or (alternatively) as a heterogeneous population extended over time; (5) that it makes sense to describe LUCA (of whatever type) as something other than a prokaryote. None of these is provably wrong, but none will ever be proved right, and all are matters of opinion, not fact. Definition-dependent, supposedly logic-driven arguments on the use of "prokaryote" seem doomed to failure by the same criteria with which they are undertaken, coupled with the vast underdetermination of any of our current theories about early cellular evolution (Vesteg and Krajcovic 2011; Forterre 2011).

Moreover, much of the debate is about the definition of terms, of which Popper wrote:

> In science, we take care that the statements we make should never *depend* on the meaning of our terms. Even where the terms are defined, we never try to derive any information from the definition, or to base any argument upon it. . . . We are always conscious that our terms are a little vague (since we have learned to use them only in practical applications) and we reach precision not by reducing their penumbra of vagueness, but rather by keeping well within it, by carefully phrasing our sentences in such a way that the possible shades of meaning of our terms do not matter. This is how we avoid quarrelling about words. (Popper 2003, pp. 21–22)

So, in a more relaxed context, but still one in which we strive to keep well within Popper's "penumbra of vagueness," is there anything sensible to say in defense of "prokaryote"?

As a phenetically based grade distinction, the prokaryote: eukaryote dichotomy resonates so strongly largely because eukaryotes seem so different from either Archaea or Bacteria, and all so much like each other, at the level of the cell. Four decades ago, the prevailing view was that eukaryotes arose from within prokaryotes by the gradual evolution of cytoskeleton and endomembrane systems—driven by selection for the acquisition of phagotrophy and followed by the establishment of the endosymbionts destined to become mitochondria and plastids (Gray and Doolittle 1982). Theory predicted that there might be some surviving primitively amitochondriate phagotrophs still out there. But so far none has been found, and most of the complexification of internal systems believed to be characteristically eukaryotic were apparently already in place in LECA. Arguably, there was very rapid, selection-directed modification of many archaeal-type genes (of the "important" kind identified by Cotton and McInerey), such that many now are hard to align with their still-prokaryotic homologs. To cell biologists working on the eukaryote side of that "great discontinuity," it is the "pro" in prokaryotes that gives meaning to the term. Especially for symbionticists like Lynn Margulis (1996), but even for more conservative theorists like John Maynard Smith and Eors Szathamary (1995), the chimeric fusion event(s) that forged eu- from prokaryotes was a profoundly important evolutionary transition.

Whether archaeal/bacterial differences are comparable in importance to the eukaryote:prokaryote distinction is indeed a matter of "comparing apples to oranges." Woese in rebutting Mayr asserted that his (Mayr's) "attempt to apply globally a parochial and subjectively defined concept of diversity serves only to reveal the futility in such an approach." But it is hard to imagine concepts of phenotypic diversity applicable to both prokaryotes and eukaryotes that are not "parochial and subjectively defined," at the very least insofar as they involve choices as to differential relevance of datasets. For the cell biologist of the school of classification Mayr called "Darwinian," focusing on the translational machinery is parochial. And for the microbiologist working with Bacteria or Archaea, although there are important differences that should inform experiments on gene structure and function, in the study of metabolism, regulation, population genetics, and ecology, this is much less the case. Without knowing in advance it would, in many instances, be hard to say whether a particular published paper on these topics had a bacterium or an archaeon as its study object.

It is not necessary to imagine that the world holds some particular number of basic types in order to understand it, or advance biological knowledge. Almost certainly "prokaryote" glosses over the many differences between Bacteria and Archaea, but "Bacteria" glosses over many differences between mycoplasmas and planctomycetes, and by what criteria can we judge one glossing-over more egregious than the other? Prokaryota is indeed not a legitimate taxonomic group by the principles of cladists, nor are Eukarya, Bacteria, and Archaea, unless we ignore or rework those principles to the point that they are no longer recognizable or useful. To avoid quarreling about words, we need first to realize that that is what we are doing.

The continued use of "prokaryote" with a small "P" seems to us justified by history and utility. If indeed it confuses students to teach them this "wrong idea," this will be because we have dumbed our science too far down. Those who come to see that "prokaryote" is neither right nor wrong and that its meaning is contextual will achieve a deeper understanding of microbiology, and of science as a human endeavor.

Acknowledgments

This chapter is dedicated to the memory of Lynn Margulis. We thank John Archibald and Andrew Roger for comments. Original work described in this manuscript was supported by the Canadian Institutes of Health Research (WFD) and by the TULA Foundation (OZ).

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Angert ER (2005) Alternatives to binary fission in bacteria. Nat Rev Microbiol 3:214–224

Ashlock PD (1974) The uses of cladistics. Ann Rev Ecol Syst 5:81–99

Baum DA (2007) Concordance trees, concordance factors, and the exploration of reticulate genealogy. Taxon 56:417–426

Bell SD, Jackson SP (2001) Mechanism and regulation of transcription in archaea. Curr Opin Microbiol 4:208–213

Bird A, Tobin E (2010) Natural kinds. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, summer 2010 edn. http://plato.stanford.edu/archives/sum2010/entries/natural-kinds/

Brown JR, Doolittle WF (1994) Root of the universal tree based on aminoacyl-tRNA gene duplications. Proc Natl Acad Sci USA 92:2441–2445

Cavalier-Smith T (2002) The neomuran origin of archarebacteria, the negibacterial root of the universal tree and bacterial megaclassification. Int J Syst Evol Microbiol 52:7–76

Cavalier-Smith T (2007) Concept of a bacterium still valid in prokaryote debate. Nature 446:257

Chatton E (1937) Titres et travaux scientifiques. Sette, Sottano, Italy

Cotton JA, McInerney JO (2010) Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of origin. Proc Natl Acad Sci USA 107:17252–17255

Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM (2008) The archaebacterial origin of eukaryotes. Proc Natl Acad Sci USA 105:20356–20361

Darnell JE, Doolittle WF (1986) Speculations on the early course of evolution. Proc Natl Acad Sci USA 83:1271–1275

Darwin CR (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life, 1st edn. John Murray, London

Di Giulio M (2011) The Last Universal Common Ancestor (LUCA) and the ancestors of Archaea and Bacteria were progenotes. J Mol Evol 72:119–126

Dolan MF, Margulis L (2007) Advances in biology reveal truth about prokaryotes. Nature 445:21

Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284:2124–2129

Doolittle WF, Brown JR (1994) Tempo, mode, the progenote, and the universal root. Proc Natl Acad Sci USA 91:6721–6728

Dupré J (1981) Natural kinds and biological taxa. Philos Rev 90:66–90

Embley TM, Martin W (2006) Eukaryote evolution, changes and challenges. Nature 440:623–630

Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. Nucl Acids Res 30:1575–1584

Esser C, Ahmadinejad N, Wiegand C, Rotte C, Sebastiani F, Gelius-Dietrich G, Henze K, Kretschmann E, Richly E, Leister D, Bryant D, Steel MA, Lockhart PJ, Penny D, Martin W (2004) A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. Mol Biol Evol 21:1643–1660

Field MC, Dacks JB (2009) First and last ancestors: reconstructing evolution of the endomembrane system with ESCRTs, vesicle coat proteins, and nuclear pore complexes. Curr Opin Cell Biol 21:4–13

Forterre P (2011) A new fusion hypothesis for the origin of Eukarya: better than previous ones, but probably also wrong. Res Microbiol 162:77–91

Foster PG, Cox CJ, Embley TM (2009) The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. Philos Trans Roy Soc B 364:2197–2207

Fox GE (2010) Origin and evolution of the ribosome. Cold Spr Harbor Perspect Biol 2010:a003483

Fox-Keller E (2010) The mirage of a space between nature and nurture. Duke University Press, Durham

Fuerst JA, Sagulenko E (2011) Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function. Nat Rev Microbiol 9:403–415

Glansdorff N, Xu Y, Labedan B (2008) The Last Universal Common AncestoR: emergence, constitution and genetic legacy of an elusive forerunner. Biol Direct 3:29

Glansdorff N, Xu Y, Labedan B (2009) The origin of life and the last universal common ancestor: do we need a change of perspective? Res Microbiol 160:522–528

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

Golinski J (1998) Making natural knowledge: Constructivism and the history of science. University of Chicago Press, Chicago

Grant V (2003) Incongruence between cladistic and taxonomic systems. Am J Bot 90:1263–1270

Gray MW, Doolittle WF (1982) Has the endosymbiont hypothesis been proven? Microbiol Rev 46:1–42

Gribaldo S, Poole AM, Daubin V, Forterre P, Brochier-Armanet C (2010) The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? Nat Rev Microbiol 8:743–752

Gupta RS (2000) The natural evolutionary relationships among prokaryotes. Crit Rev Microbiol 26:111–131

Guy L, Ettema TJG (2011) The archaeal "TACK" superphylum and the origin of eukaryotes. Trends Microbiol 19:580–587

Hacking I (1991) A tradition of natural kinds. Philos Stud 61:109–126

Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T (1989) Evolutionary relationship of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci USA 86:9355–9459

Jain R, Rivera MC, Lake JA (1999) Horizontal gene transfer among genomes: the complexity hypothesis. Proc Natl Acad Sci USA 96:3801–3806

Koonin EV (2009) On the origin of cells and viruses: primordial virus world scenario. Ann NY Acad Sci 1178:47–64

Lake JA, Henderson E, Oakes M, Clark MW (1984) Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proc Natl Acad Sci USA 81:3786–3790

Lovejoy AO (1936) The great chain of being. Harvard University Press, Cambridge

Lukes J, Archibald JM, Keeling PJ, Doolittle WF, Gray MW (2011) How a neutral evolutionary ratchet can build cellular complexity. IUBMB Life 63:528–537

Makarova KS, Yutin N, Bell SD, Koonin EV (2010) Evolution of diverse cell division and vesicle formations systems in Archaea. Nat Rev Microbiol 8:731–741

Margulis L (1996) Archaeal-eubacterial mergers in the origin of Eukarya: phylogenetic classification of life. Proc Natl Acad Sci USA 93:1071–1076

Martin WF (2011) Early evolution without a tree of life. Biol Direct 6:36

Martin W, Koonin E (2006) A positive definition of prokaryotes. Nature 442:24

Martin WF, Russell MJ (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos Trans R Soc Lond B Biol 358:59–63

Maynard Smith J, Szathmary E (1995) The major transitions in evolution. Oxford University Press, Oxford

Mayr E (1982) The growth of biological thought. Harvard University Press, Cambridge, MA

Mayr E (1990) A natural system of organisms. Nature 348:491

Mayr E (1991) More natural classification. Nature 353:122

Mayr E (1998) Two empires or three? Proc Natl Acad Sci USA 95:9720–9723

McInerney JO, Martin WF, Koonin EV, Allen JF, Galperin MY, Lane N, Archibald JM, Embley TM (2011) Plactomycetes and eukaryotes: a case of analogy not homology. Bioessays 33:810–817

McOuat G (2009) The origins of "natural kinds": keeping "essentialism" at bay in the Age of Reform. Intell Hist Rev 19:211–230

Mindell DP (1992) Phylogenetic consequences of symbioses: Eukarya and eubacteria are not monophyletic taxa. Biosystems 27:53–62

Mishler BD, Theriot EC (2000) The phylogenetic species concept (*sensu* Mishler and Theriot): monophyly, apomorphy and phylogenetic species concepts. In: Wheeler QD, Meier R (eds) Species concepts and phylogenetic theory. Columbia University Press, New York

O'Brien TW (2002) Evolution of a protein-rich mitochondrial ribosome: implications for human genetic disease. Gene 286:73–79

O'Hara, Robert J (1997) Population thinking and tree thinking in systematics. Zool Scr 26(4):323–329

O'Hara RJ (1998) Population thinking and tree thinking in systematics. Zoologica Scripta 26:323–329

O'Malley MA (2010) Ersnt Mayr, the tree of life, and philosophy of biology. Biol Philos 25:529–552

Pace NR (2006) Time for a change. Nature 441:289

Pace NR (2009) Rebuttal: the modern concept of the prokaryote. J Bacteriol 191:2006–2007

Pace NR, Sapp J, Goldenfeld N (2012) Phylogeny and beyond: scientific, historical, and conceptual significance of the first tree of life. Proc Natl Acad Sci USA. Early edition on line

Payankaulam S, Li ML, Arnosti DN (2010) Transcriptional repression: conserved and evolved features. Curr Biol 20:R764–R771

Peeters E, Charlier D (2010) The Lrp family of transcription regulators in archaea. Archaea 2010:750457

Penny D, Poole A (1999) The nature of the last universal common ancestor. Curr Opin Genet Dev 9:672–677

Pisani D, Cotton JA, McInereny JO (2007) Supertrees disentangle the chimerical origin of eukaryotic genomes. Mol Biol Evol 24:1752–1760

Popper K (2003) The open society and its enemies, vol 2. Routledge, London (first published 1945)

Poptsova MS, Gogarten JP (2007) BranchClust: a phylogenetic algorithm for selecting gene families. BMC Bioinformatics 8:120

Roberts E, Sethi A, Montoya J, Woese CR, Luthey-Schluten Z (2008) Molecular signatures of ribosomal evolution. Proc Natl Acad Sci USA 105:13953–13958

Sapp J (2006) Two faces of the prokaryote concept. Int Microbiol 9:163–172

Sapp J (2009a) The prokaryote-eukaryote dichotomy: meanings and mythology. Microbiol Mol Biol Rev 69:292–305

Sapp J (2009b) The new foundations of evolution. Oxford University Press, Oxford

Sneath PHA (1975) Cladistic representation of reticulate evolution. Syst Zool 24:360–368

Spath SB (1999) C.B. van Niel and the culture of microbiology, 1920–1965. PhD thesis, History, University of California, Berkeley

Stanier RY, Doudoroff M, Adelberg EA (1963) The microbial world. Prentice-Hall, Engelwood Cliffs, NJ

Stanier RY, van Niel CB (1962) The concept of a bacterium. Arch Microbiol 42:17–35

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41

Vesteg M, Krajcovic J (2011) The falsifiability of the models for the origin of eukaryotes. Curr Genet 57:367–390

Vetsigian K, Woese D, Goldenfeld N (2006) Collective evolution and the genetic code. Proc Natl Acad Sci USA 103:10696–10701

Walsh DA, Doolittle WF (2005) The real 'domains' of life. Curr Biol 15: R237–R240

Werner F, Grohmann D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. Nat Rev Microbiol 9:85–98

Whitman WB (2009) The modern concept of the prokaryote. J Bacteriol 191:2000–2005

Winsor MP (2006) Linnaeus's biology was not essentialist. Ann Missouri Botanical Garden 93:2–7

Woese CR (1998a) Default taxonomy: Ernst Mayr's view of the microbial world. Proc Natl Acad Sci USA 95:11043–11046

Woese C (1998b) The universal ancestor. Proc Natl Acad Sci USA 95:6854–6859

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. Proc Natl Acad Sci USA 87:4576–4579

Zhaxybayeva O, Gogarten JP (2004) Cladogenesis, coalescence and the evolution of three domains of life. Trends Genet 20:182–187