Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 6854–6859, June 1998
Evolution

The universal ancestor

(progenote/lateral gene transfer/genetic annealing/evolutionary temperature/communal ancestor)

CARL WOESE*

Department of Microbiology, University of Illinois at Urbana-Champaign, B103 Chemical and Life Sciences Laboratory, MC-110, 601 South Goodwin Avenue, Urbana, IL 61801

Contributed by Carl R. Woese, April 3, 1998

ABSTRACT **A genetic annealing model for the universal ancestor of all extant life is presented; the name of the model derives from its resemblance to physical annealing. The scenario pictured starts when "genetic temperatures" were very high, cellular entities (progenotes) were very simple, and information processing systems were inaccurate. Initially, both mutation rate and lateral gene transfer levels were elevated. The latter was pandemic and pervasive to the extent that it, not vertical inheritance, defined the evolutionary dynamic. As increasingly complex and precise biological structures and processes evolved, both the mutation rate and the scope and level of lateral gene transfer, i.e., evolutionary temperature, dropped, and the evolutionary dynamic gradually became that characteristic of modern cells. The various subsystems of the cell "crystallized," i.e., became refractory to lateral gene transfer, at different stages of "cooling," with the translation apparatus probably crystallizing first. 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.**

BACKGROUND

Biologists have long subscribed to the powerful, unifying idea that all life on Earth arose from a common ancestor (1). Nothing concrete could be said about the nature of this ancestor initially, but it was intuitively assumed to be simple, often likened to a prokaryote, and generally held to have had little or no intermediary metabolism (2). Only when biology could be defined on the level of molecular sequences would it become possible to seriously question the nature of this ancestor.

The unrooted universal phylogenetic tree that emerged from ribosomal RNA (rRNA) sequence comparisons provided the first glimpse of our ultimate ancestor, albeit an indirect one (3, 4). Whatever it was, this cryptic entity had spawned three remarkably different primary groupings of organisms (domains)—the Archaea, the Bacteria, and the Eucarya–and these necessarily reflected the ancestor's nature. Phylogenies derived from the few other molecules that then had been sequenced confirmed the three predicted groupings, and concurrent biochemical characterizations further developed their uniqueness (5–12). But, from this first universal tree, one could infer only that the ancestor was some ill-defined "urstoff" from which three primary lines of descent somehow arose (3, 13).

When it proved possible to root the tree, by using the Schwartz–Dayhoff paralogous gene outgroup method (14–16), the ancestor became a node on the tree, implying that it was a specific entity. This rooted tree also unexpectedly revealed the Archaea to be specific relatives of the eukaryotes. If prokaryotes (Archaea and Bacteria) were on both sides of the primary phylogenetic divide, then "prokaryote" was not a phylogenetically meaningful taxon. In addition, given the fundamental molecular differences between Archaea and Bacteria, it made no sense to call the universal ancestor a "prokaryote." What then was this universal ancestor?

A discrete picture of the ancestor began to emerge only when many more sequences representing all three phylogenetic domains became available. These sequences could be seen as putting phenotypic flesh on an ancestral phylogenetic skeleton. Yet that task has turned out to be anything but straightforward. Indeed, it would seem to require disarticulating the skeleton. No consistent organismal phylogeny has emerged from the many individual protein phylogenies so far produced.

Phylogenetic incongruities can be seen everywhere in the universal tree, from its root to the major branchings within and among the various taxa to the makeup of the primary groupings themselves. Yet there is no consistent alternative to the rRNA phylogeny, and that phylogeny is supported by a number of fundamental genes. The aminoacyl–tRNA synthetases (aaRSs) epitomize this confused situation (17, 18). For example, it is common to see archaeal versions of some of the aaRSs scattered throughout the Bacteria (17) (C.W., unpublished data). The aaRSs can in principle be used to root the universal tree (because some of them obviously reflect common ancestral gene duplications). Yet different (related) aaRSs root that tree differently: the ileRS tree roots (using the valRSs) canonically; i.e., the Archaea and eukaryotes are sister groups (19). The valRS tree, however, roots on the archaeal branch, which makes sister groups of the Bacteria and eukaryotes (C.W., unpublished data). Exceptions to the topology of the rRNA tree such as these are sufficiently frequent and statistically solid that they can be neither overlooked nor trivially dismissed on methodological grounds. Collectively, these conflicting gene histories are so convoluted that lateral gene transfer is their only reasonable explanation (18).

A concept of the universal ancestor turns on more than phylogenetic trees, however. The Archaea and Bacteria share a large number of metabolic genes that are not found in eukaryotes (18, 20). If these two "prokaryotic" groups span the primary phylogenetic divide and their genes are vertically (genealogically) inherited, then the universal ancestor must

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: rRNA, ribosomal RNA; RS, tRNA synthetase; aaRS, aminoacyl-tRNA synthetase.

*To whom reprint requests should be addressed. e-mail: carl@ninja.life.uiuc.edu.

have had all of these genes, these many functions: This distribution of genes would make the ancestor a prototroph with a complete tricarboxylic acid cycle, polysaccharide metabolism, both sulfur oxidation and reduction, and nitrogen fixation; it was motile by means of flagella; it had a regulated cell cycle, and more. This is not the simple ancestor, limited in metabolic capabilities, that biologists originally intuited. That ancestor can explain neither this broad distribution of diverse metabolic functions nor the early origin of autotrophy implied by this distribution. The ancestor that this broad spread of metabolic genes demands is totipotent (21), a genetically rich and complex entity, as rich and complex as any modern cell—seemingly more so.

Yet the totipotent ancestor also fails: it cannot explain the manner of the ancestor's evolution, i.e., how it became so miraculously complex in so short a time and just as rapidly gave rise to the ancestors of the three primary lines of descent. All of this apparently happened in far less than 1 billion years, whereas evolution within each of the three primary lines of descent has been going on for over 3 billion years now with outcomes that don't even begin to compare with the spectacular ones associated with the ancestor and its original offspring (4)—yet experience teaches that complex, integrated structures change more slowly than do simple ones. Moreover, the totipotent ancestor associates physiologies that have not been observed together in any modern lineage and asks that all of this come about through vertical inheritance. Thus, we are left with no consistent and satisfactory picture of the universal ancestor. It is time to question underlying assumptions.

The Pivotal Assumption. Most theories of early evolution tacitly assume that organismal lineages, organismal genealogies, have always existed and extend into the stage of the universal ancestor. Eukaryotes, of course, contain organellar genes, whose heritages are not those of nuclear genes in general. Laterally transferred genes are seen in prokaryotes as well. Strictly speaking both eukaryotes and prokaryotes are of mixed heritage. Yet, we still speak of eukaryotic and prokaryotic "lineages" (and for good reason) because in both cases the vast majority of their genes presumably share a common history. If and only if this assumption holds, however, can we speak of organismal lineages and corresponding phylogenetic trees. But the assumption automatically makes the universal ancestor an organism that itself had a lineage, a discrete genealogy.

The further back in evolutionary time we look, the more the notion of an "organismal lineage"—indeed, the very definition of "organism" itself—comes into question. It is time to release this notion of organismal lineages altogether and see where that leaves us. Let molecular phylogenetic trees represent exactly what they in the first instance do represent, histories of individual genes or gene groupings. When do individual gene histories define an organismal history, an organismal lineage? Did organismal lineages even exist at the time of the universal ancestor? If not, then what exactly was this ancestor, and what was the nature of the evolutionary process that formed it?

THE GENETIC ANNEALING MODEL

A very different picture of the universal ancestor comes to light when the notion of organismal lineages is released, a picture that flows naturally from a consideration of what the evolutionary dynamic might be when cells are very primitive. I have been developing parts of this idea in various publications over the last three decades (4, 22–26). Now, in the context of far greater amounts of molecular sequence data, a synthesis is emerging. The primitive evolutionary dynamic I envision bears a superficial resemblance to the physical annealing process, hence, its name.

First consider the analogy: a physical annealing system starts at a high enough temperature that structures cannot form and then proceeds to slowly cool. In this quasi-stable condition, various combinations of the system's elements form, dissociate, and reform in new ways, with only the most stable and structured of these combinations initially persisting, i.e., "crystallizing." As the temperature continues to drop, less stable structures begin to form, to crystallize, and many of the preexisting ones add new features, becoming more elaborate. In the evolutionary counterpart of physical annealing, the elements of the system are primitive cells, mobile genetic elements, and so on, and physical temperature becomes "evolutionary temperature," the evolutionary tempo. The evolutionary analog of "crystallization" is emergence of new structures, new cellular subsystems that are refractory to major evolutionary change (see below). The analogy between physical cooling and the drop in evolutionary temperature is somewhat inexact, as we shall see. And although the outcomes of a physical annealing process are highly circumscribed if not certain, the evolutionary world is open-ended to an extreme.

Primitive Cells: Progenotes. The scenario about to unfold starts at the point when the translation mechanism first came into being. (How it arose does not concern us here). It is assumed that cells existed at this time but were very different from modern cells, different enough that they should not be looked at as organisms (see below). The properties of the rudimentary translation mechanism severely limited these cells in both their nature and evolutionary potential. The rudimentary translation mechanism was far simpler than the complex modern one and, as a consequence, was far less accurate (26); codon recognition and reading frame movement (procession) were both so inaccurate that most, if not all, modern types of proteins could not be produced. At this stage, only small proteins could evolve—along with any larger, imprecisely translated ones (called "statistical proteins") that the primitive cell was able to produce and use (26, 27). Entities in which translation had not yet developed to the point that proteins of the modern type could arise have been termed "progenotes," and the era during which these were the most advanced forms of life, the "progenote era" (26).

If modern large proteins could not be produced by progenotes, then a modern type of genome replication/repair mechanism did not exist. As with translation, a rudimentary mechanism implies a less accurate one (26), and the resulting high mutation rates necessitated small genomes.

The structure of these genomes must reflect the primitive evolutionary dynamic in general. Therefore, I see the progenote genome as organized rather like the macronucleus of some ciliates today (25): it comprised many small linear chromosomes (mini-chromosomes), each present in multiple copies. Each chromosome was "operonally" organized, that is, functionally or structurally related genes were grouped together. The individual chromosomes were "semi-autonomous" in the sense that they more resembled mobile genetic elements than typical modern chromosomes. Cell division occurred in the simplest way possible, by a physical pinching of the cell into two approximately equal halves.

Small chromosomes are demanded because, when mutation rates are high, only these stand a chance of being replicated without a crippling number of mutations. A linear (small) chromosome makes both replication and transcription simple from a topological perspective (topoisomerases don't seem to be needed). Chromosome multiplicity means genetic redundancy, which serves to ensure functionality when one or more copies of a gene are knocked out. Operonal organization is selected for by both the random mode of chromosome segregation (at cell division) and more strongly yet by lateral gene transfer (28)—there is little or no benefit in inheriting only part of a new metabolic pathway. And mobile genetic elements are well suited for lateral transfer as well.

Upon cell division, the mini-chromosomes distribute randomly between the daughter cells. This fact would lessen the

mutational burden imposed by high mutation rates in the sense that the daughter cell with the better balance of functional copies of important genes has a selective advantage. If replication errors could be directly detected (e.g., as mispairings), a more direct way to eliminate them seems possible, through simple destruction of the mini-chromosome in question, say, by nuclease cleavage (25).

Small primitive genomes with low genetic capacity and imprecision in both translation and genome replication imply a primitive cell that was rudimentary in every respect (26). The progenote probably had no cell wall (see below) (13). Its subsystems were generally less complex and hierarchically organized and the cell itself was less integrated than are cells today. The states of that cell were fewer, simpler, and imprecisely defined and controlled (26). The progenote was more or less a bag of semi-autonomous genetic elements (the mini-chromosomes). These elements would come and go, especially on an evolutionary time scale. Higher level organization, among the mini-chromosomes and throughout the cell, was minimal.

Evolutionary Temperature. Macroscopic evolutionists recognized long ago a relationship between the "tempo" (rate) of evolution and what they called its "mode" (a measure of the outcomes): the more rapid the former, the more unusual and varied the latter (29, 30). When microbial evolution finally came into the picture, a similar (and seemingly related) phenomenon was encountered on the molecular level (4), suggesting that this tempo/mode relationship was a fundamental manifestation of the evolutionary process. Because of high mutation rates and other factors (below), the progenote era is seen as one of very high evolutionary tempo.

Evolutionary tempo, i.e., "temperature," is defined here as a composite of the two processes critical to evolution: (*i*) mutation and the fields of variants that result, and (*ii*) lateral gene transfer, including its frequency and quality. A lineage's field of variants, the anlage for evolutionary change, is a strong function of mutation rate. Multi-site variants, the more useful, creative ones, obviously occur as higher order functions of this rate. These variants will disproportionately increase and become more varied in kind as mutation rate increases, and that increase will, in effect, qualitatively change the field of variants, changing the mode of evolution (4).

The field of variants in which progenotes evolve may be even richer than that so far implied. Cell lines are capable of going into error catastrophe, a state in which their mutation rates increase out of control (31, 32), and the line replicates itself into extinction in short order. These short-lived, error-prone cell lines take on special significance in a world of lateral gene transfer (below). In this context, they become "super-mutator" strains for the population as a whole. The variant genes they produce, ones that viable cell lines cannot afford to produce, add great richness to the delocalized field of variants from which all progenotes can draw.

The primitive lateral gene transfer envisioned is very unlike that seen today. It effectively involves all entities existing at the time and all of their genes, and transfers, like mutations, occur at very high frequencies. The reason why all cellular entities are potential recipients and all genes potentially transferable is that progenotes in essence comprised what would now be called the "essential functions," and primitive evolution (as measured by its outcomes) was concerned with the development and refinement of these. All functions of this sort and their refinements could be globally exchanged. The high frequencies of lateral transfer reflected the simplicity of the progenote's genetic mechanisms and the lack of barriers to lateral exchange—in this primitive context any lineages evolving barriers to acquisition or expression of foreign genes would be left behind in the evolutionary progression toward modern life. Lateral gene transfer of this kind and intensity would not only contribute significantly to but also would completely dominate the primitive evolutionary dynamic.

The Communal Ancestor. Progenotes were very unlike modern cells. Their component parts had different ancestries, and the complexion of their componentry changed drastically over time. All possessed the machinery for gene expression and genome replication and at least some rudimentary capacity for cell division. But even these common functions had no genealogical continuity, for they too were subject to the confusion of lateral gene transfer. Progenotes are cell lines without pedigrees, without long-term genetic histories. With no organismal history, no individuality or "self-recognition," progenotes are not "organisms" in any conventional sense.

Their small genomes require progenotes to be metabolically simple, minimal. However, different progenotes could have differed metabolically. The collectively genetic complement of the progenote population could have been far greater than that of any individual cell, indeed totipotent in the above sense (13, 22). The fact that innovations could easily spread through the population by lateral transfer gave the progenote community enormous evolutionary potential; each cell line was the potential recipient of any innovation that occurred within the entire diverse population.

There are different ways of looking at such a community of progenotes. On the one hand, it could have been the loose-knit evolutionary (genetic) community just discussed. On the other, it could have been more like a modern bacterial consortium, with cells cross-feeding one another not only genetically but also metabolically. Cell–cell contacts would have facilitated both processes. In both views of the community, the latter in particular, it is not individual cell lines but the community of progenotes as a whole that survives and evolves. It was such a community of progenotes, not any specific organism, any single lineage, that was our universal ancestor—a genetically rich, distributed, communal ancestor. It was also this loose-knit biological unit that ultimately evolved to a stage in which it somehow pulled apart into two, then three communities, isolated by the fact that they could no longer communicate laterally with one another in an unrestricted way. Each had become sufficiently complex and idiosyncratic that only some genes, some subsystems, could be usefully transferred laterally. Each of these three self-defining communities then further congealed, giving rise to what we perceive as the three "primary lines of descent."

Translation Improves: Progenotes Become Genotes. At these early stages of life, everything turned upon the evolution of translation. Each slight improvement in that process, each increase in its accuracy, would have permitted a new generation of proteins to emerge (26). These new proteins, in turn, refined and developed the metabolic pathways and generally improved the cell, which then set the stage for a further round of improvement in translation. In this way, wave after wave of innovation occurred, each triggered by a refinement in translation and spread throughout the community by lateral gene transfer. This iterative, bootstrapping evolution continued until the accuracy of translation reached a level where it no longer prevented the evolution of the types of proteins we see today. The evolutionary dynamic then ceased to be constrained by imprecise translation, and progenotes, by definition, became genotes (26). This transition did not mean that translation had stopped evolving, nor did it mean that the initial genotes were modern types of cells. That latter development required many more innovations and refinements.

Cooling. Evolutionary temperature is postulated to drop gradually during the primitive evolutionary process. This cooling, however, does not bring about the crystallization of structures as in physical annealing: evolutionary cooling occurs as a result of crystallization. All structures in the progenote cell, all cellular subsystems, are initially simple, as are their relationships and as is the cell itself. As progenotes evolve

(above), structures of increasing complexity emerge, relationships among them become more intricate, and the cell itself becomes more integrated, more highly organized. In the process, individual cells (cell lines) become increasingly dissimilar, increasingly idiosyncratic. In other words, the biological specificity of the system increases in every respect.

The more complex a subsystem becomes, the harder it is to find a foreign part compatible with it, and the few that are tend to come from cells that have related subsystems. The more a subsystem becomes integrated into the fabric of the cell, the harder it becomes to replace it *in toto*. There comes a stage, then, when the subsystem can no longer change through lateral gene transfer: all changes must come from within the cell line, through gene duplications and mutations. At this point, the subsystem has crystallized, evolving essentially through vertical inheritance.

Crystallizing. The annealing scenario predicts that different subsystems of the cell will crystallize (become more or less refractory to lateral transfer) at different evolutionary stages. This point will be reached when (as just stated) foreign parts are no longer compatible with the subsystem, and it becomes firmly integrated into the fabric of the cell.

I would argue that translation was among the first, if not the first, of the cellular subsystems to crystallize: The fact that translation is an RNA-based mechanism suggests antiquity. The fact that it is complex and its key components tend to be universal argues for an early consolidation as well. And, as the leading edge of the early evolutionary waves, translation would have refined at each step before the other cellular systems did.

Not all of translation's components are universal in distribution, indicating that the mechanism continued to refine (in yet to be understood ways) after its core had crystallized and the stage of the universal ancestor had passed. These later refinements do not appear to involve lateral gene transfer although subtle forms of that transfer (involving relatively closely related organisms) cannot be ruled out. Immunity to lateral transfer would be expected for a mechanism so complex (idiosyncratic) and tightly integrated, i.e., one that had crystallized.

The aaRSs are telling exceptions to the vertical inheritance that characterizes the other translational components. These enzymes are a study in lateral transfer. For example, several different versions of a given synthetase often occur within the Bacteria alone. (Insufficient data prevent assessing this for the Archaea and eukaryotes.) For the different bacterial versions of a given synthetase: (*i*) more than one of them can be simultaneously present in the same organism; (*ii*) the taxonomic makeup of these synthetase subgroupings are individually idiosyncratic (none seems to conform to established phylogenetic pattern, or to agree with the others); and (*iii*) some bacterial versions of a given enzyme are more related to the archaeal and/or eukaryotic versions than they are to the other bacterial versions (17) (C.W., unpublished data). These phenomena are all indicative of lateral transfer of the synthetase genes throughout the evolution of the Bacteria.

It is obvious why the aaRSs are so evolutionarily migratory (17, 18). They are "modular": they occur free in the cell, unassociated with the ribosome; each synthetase type interacts with only one or a few tRNAs; and their functions are universal. This all adds up to a capacity to function in a great variety of foreign cellular environments. Lateral movement and diversification should characterize other modular elements as well, and there is mounting evidence to support this: sulfate reduction, for example, appears to have been laterally transferred between Bacteria and Archaea, if not within the bacteria as well (20, 33).

Transcription, too, seems to have crystallized at an early stage, although it is not known whether this stage was before or after translation. What can be said is that the first transcription mechanism to achieve genealogical coherence was only a rudimentary one. Substantial differences in the mechanism separate the Bacteria on the one hand from the Archaea and eukaryotes on the other: all versions of the polymerase possess the core subunits, i.e., α, β, and β' in bacterial terminology, but these subunits differ substantially in sequence, particularly in the case of α, in which an obvious deep structural divergence strongly distinguishes the bacterial and archaeal/eukaryotic versions (34). The Archaea and eukaryotes have a number of additional common subunits that are not seen in the bacterial mechanism (34). As was the case with translation, there is little evidence to suggest that lateral transfer was involved in the evolution of transcription; all of the components of the apparatus seem to provide the same genealogical pattern. Of course, the question of phylogenetically local transfers remains open.

Genome replication presents a different picture. No genome replication system is universally distributed; the bacterial mechanism bears no specific relationship to the one that is basically common between the Archaea and eukaryotes (21). Such a universal mechanism probably existed early on (as is suggested by the general homologies among various types of DNA polymerases spread across the phylogenetic spectrum), but that mechanism must have been too primitive to be simply refined into those we see today. It stands to reason that the evolution of (a particular) genome organization goes hand in hand with that of the corresponding mechanism to replicate it. Therefore, modern genomes appear to have arisen only after the primary lines of descent were established (21), and the evolution leading to their replication mechanisms involved major innovations, innovations that did (could) not spread globally.

Cellular evolution, the emergence of modern cells, seems a protracted process with a somewhat ill-defined ending. The evolution of modern genome structure and genome replication mechanisms appears the last great innovation in the evolution of the cell, and so, it may have marked the beginning of the end of that process. For genomes to reach the size of modern genomes, the mutation rate has to be low, in the range of one error per 1 billion base pairs read (32). This rate lowers the evolutionary temperature (at least the mutational aspect thereof) to modern levels. And the implied complexity and specialization of cells and their subsystems at this point should restrict lateral transfer significantly.

The Universal Phylogenetic Tree. By now, it is obvious that what we have come to call the universal phylogenetic tree is no conventional organismal tree. Its primary branchings reflect the common history of central components of the ribosome, components of the transcription apparatus, and a few other genes. But that is all. In its deep branches, this tree is merely a gene tree. Genuine organisms (self-replicating entities that have true individuality and a history of their own) did not exist at the time the tree started to form. The tree arose in a communal universal ancestor, an "entity" that had a physical history but not a genealogical one. This tree became an organismal tree only as it grew, only as its more superficial branches emerged. By the time these formed, many more functions had crystallized and so, had come to have discernible histories; and these histories coincided with those of the ribosomal components and the like—but only *after* the point of their crystallization.

An interesting question is whether the universal tree had become an organismal tree by the time the three primary lines of descent began to form and branch. I think not. The cellular design commitments implied by the existence and vertical evolution of the bacterial, archaeal, and eukaryotic ribosome types should preclude many of the evolutionary innovations that occurred in one of the primary organismal groups from being successfully transferred laterally to one or both the others; but, in that the genotes of that day were less sophisticated than modern cells (and the evolutionary temperature

was still elevated), more of these innovations might still transfer globally than would later be the case. I picture the ancestors of each of the primary lines of descent as being themselves to some extent communal, but in a much more local, restricted sense than that which holds for the universal ancestor. Were this true, the major lineages within each organismal domain would not sort out cleanly along phylogenetic lines: there would be many conflicting gene histories for the deep branchings within each domain. [In that the translation apparatus is one of the most evolutionarily stable of the subsystems in the cell, rRNA has to be an especially reliable indicator of the true organismal branching patterns (35).]

The Problem of Shared Metabolic Genes. The genetic annealing model does not (now) account satisfactorily for the large number of metabolic genes that are shared by the Archaea and Bacteria but not found in the eukaryotes. It does, however, suggest a new way of looking at the problem. These genes are shared, not because of vertical inheritance but because of lateral gene transfer. Metabolic functions are among the most modular in the cell, and so, their genes are expected to travel laterally, even today. Many cases now are known in which a bacterial metabolic gene occurs in one or a few Archaea or vice versa. Cases of seemingly lateral transfer within the Bacteria or within the Archaea also occur, even more frequently. However, sporadic lateral transfer of a bacterial or an archaeal metabolic gene is one thing, transfers that result in a broad, if not universal, distribution of a metabolic gene within both the Archaea and the Bacteria may be another. It would help to have many more genomic sequences, so that the distributions of these genes can be defined in some detail and their phylogenetic relationships can be determined. Then we would be in a much better position to interpret their "universal" distribution among the Bacteria and Archaea.

The sequence similarities among the various versions of a given metabolic enzyme may be of some help in understanding their organismal distribution. It would appear that, in most cases, specific archaeal and bacterial versions of the genes in question can be recognized. This means that were lateral transfer responsible for their organismal distribution, the gene transfers effectively ceased before the first branchings occurred within either of the domains.

Although the bacterial and archaeal versions of the enzyme are distinguishable, the sequence distinctions between the two versions tend to be relatively minor—not like the profound differences that separate the bacterial and archaeal versions of various components of the translation or transcription machineries. This lack of genuinely telling differences between the bacterial and archaeal versions is exactly what would happen for genes that transferred laterally well after the ribosome had crystallized but had themselves crystallized before the initial branchings within each of the primary lines of descent had occurred. (A tree based on a molecule with these lateral transfer characteristics would be congruent with the rRNA tree, but, unlike the rRNA tree, it would be "bushy" at its base.)

However, none of this addresses the absence of these genes in eukaryotes. Given that metabolic genes tend to be laterally mobile and that the eukaryotes engage in lateral gene transfers, especially (but not exclusively) through endosymbioses, it is reasonable to expect that the eukaryotes had no opportunity to sample the genes in question. Thus, the lack of these metabolic genes in eukaryotes seems more related to the nature of the early eukaryotic cell than to any specific ancestral relationship between Archaea and Bacteria. When the genomes of some of the deeply branching eukaryotes have been sequenced, the perspective to resolve this problem may exist.

CONCLUSION

The Universal Ancestor. The genetic annealing model is an attempt to develop a consistent general picture of the universal ancestor, and it almost succeeds at this. The ancestor cannot have been a particular organism, a single organismal lineage. It was communal (13, 22), a loosely knit, diverse conglomeration of primitive cells that evolved as a unit, and it eventually developed to a stage where it broke into several distinct communities, which in their turn become the three primary lines of descent. The primary lines, however, were not conventional lineages. Each represented a progressive consolidation of the corresponding community into a smaller number of more complex cell types, which ultimately developed into the ancestor(s) of that organismal domain. The universal ancestor is not an entity, not a thing. It is a process characteristic of a particular evolutionary stage.

Lateral Gene Transfer. Lateral gene transfer, which has long been recognized as a secondary evolutionary mechanism, becomes primary in this primitive evolutionary dynamic. It is through lateral transfer, not vertical inheritance, that systems primarily evolve at the progenote stage. As a result of genetic mixing, organismal lineages, consensus histories of an organism's genes, did not exist, although short-term "cell lines" necessarily did. The universal ancestor does have an evolutionary history, but that history is physical, not genealogical.

Evolution in the progenote era can be seen as occurring on the subcellular level, although it actually happens in the context of (primitive) cells. The distinction here is that, in the modern world, evolutionary innovation tends to become established through selection acting on organisms, whereas in a world dominated by lateral gene transfer, an innovation takes over by direct "invasion." The organism (organismal lineage) that carries the innovation also brings with it all its other idiosyncrasies, which are potential determinants of the future evolutionary course. The innovation established through lateral transfer, however, becomes stripped of extraneous genetic baggage by that process. Evolution at the subcellular level can be viewed as a bridge between modern organismal evolution and the much earlier evolution that involved "organic" chemicals in an abiotic world.

Allow me one final word about lateral gene transfer: The genetic annealing model sees two aspects to genetic temperature, mutation rate and the level of lateral gene transfer, which loosely covary. The question is whether or not this connection held only in the past, only in a world of progenotes. It is now clear that lateral transfer is far more widespread than had previously been appreciated (28), and that episodes of rapid evolution (high evolutionary temperature) have been common throughout evolution (4). The rRNA signature of this increased evolutionary temperature is unusually long ancestral branches on an rRNA tree (4), branches that are sparsely populated taxonomically (have few side branches). The length of these branches in part reflects unusual variations in the underlying rRNA sequences (4). What is seen in terms of rRNA sequence in these cases is presumably mirrored at the protein sequence level: (vertically inherited) protein genes would be more highly diverged than are their relatives in slowly evolving sister lineages (4). Now we add to this conjecture that the genomes resulting from episodes of rapid evolution will contain an abnormally high proportion of foreign genes. Genome sequences will soon be available in sufficient number to properly test whether the tempo/mode relationship (rapid evolution) invariably links increased mutation rate and increased levels of lateral gene transfer or vice versa.

I am indebted to Gary Olsen, Charles Kurland, and Ross Overbeek, for critical suggestions regarding details of the argument, to Norman Pace and David Graham for important suggestions as to making the final manuscript more readable, and to Claudia Reich for editorial

suggestions. I am most indebted to Gary Olsen for the amount of time, trenchant analysis, and care he put into helping me present this thesis in an understandable manner. The author is supported in part by a grant from the National Aeronautics and Space Administration.

1. Darwin, C. C. (1859) *The Origin of the Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life* (John Murray, London).
2. van Niel, C. B. (1946) *Cold Spring Harbor Symp. Quant. Biol.* **11,** 285–301.
3. Woese, C. R. & Fox, G. E. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5088–5090.
4. Woese, C. R. (1987) *Microbiol. Rev.* **51,** 221–271.
5. Langworthy, T. A. (1985) in *The Bacteria* (Academic, New York), Vol. 8, pp. 459–497.
6. Gupta, R. & Woese, C. R. (1980) *Curr. Microbiol.* **4,** 245–249.
7. McCloskey, J. A. (1986) *Syst. Appl. Microbiol.* **7,** 246–252.
8. Ramirez, C., Shimmin, L. C., Newton, C. H., Matheson, A. T. & Dennis, P. P. (1989) *Can. J. Microbiol.* **35,** 234–244.
9. Matheson, A. T., Louie, K. A. & Henderson, G. N. (1986) *Syst. Appl. Microbiol.* **7,** 147–150.
10. Prangishvilli, D., Zillig, W., Gierl, A., Biesert, L. & Holz, I. (1982) *Eur. J. Biochem.* **122,** 471–477.
11. Lechner, K. & Böck, A. (1987) *Mol. Gen. Genet.* **208,** 523–528.
12. Kandler, O. (1982) *Archaebacteria* (Gustav Fischer, Stuttgart, Germany).
13. Kandler, O. (1994) *J. Biol. Phys.* **20,** 165–169.
14. Schwartz, R. M. & Dayhoff, M. O. (1978) *Science* **199,** 395–403.
15. Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, R. J., Date, T., Oshima, T., *et al*. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6661–6665.
16. Iwabe, N., Kuma, K.-i., Hasegawa, M., Osawa, S. & Miyata, T. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 9355–9359.
17. Shiba, K., Motegi, H. & Schimmel, P. (1997) *Trends Biochem. Sci.* **22,** 453–457.
18. Brown, J. R. & Doolittle, W. F. (1997) *Microbiol. Mol. Biol. Rev.* **61,** 456–502.
19. Brown, J. R. & Doolittle, W. F. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 2441–2445.
20. Klenk, H.-P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., *et al*. (1997) *Nature (London)* **390,** 364–370.
21. Olsen, G. J. & Woese, C. R. (1996) *Trends Genet.* **12,** 377–379.
22. Woese, C. R. (1982) *Zentralbl. Bakteriol. Mikrobiol. Hyg. Ser. C.* **3,** 1–17.
23. Woese, C. R. (1967) *The Genetic Code: The Molecular Basis for Genetic Expression* (Harper and Row, New York).
24. Woese, C. R. (1972) in *Exobiology*, ed. Ponnamperuma, C. (North–Holland, Amsterdam), pp. 301–341.
25. Woese, C. R. (1983) in *Evolution from Molecules to Men*, ed. Bendall, D. S. (Cambridge Univ. Press, Cambridge, U.K.), pp. 209–233.
26. Woese, C. R. & Fox, G. E. (1977) *J. Mol. Evol.* **10,** 1–6.
27. Woese, C. R. (1965) *Proc. Natl. Acad. Sci. USA* **54,** 1546–1552.
28. Lawrence, J. G. & Roth, J. R. (1996) *Genetics* **143,** 1843–1860.
29. Mayr, E. (1942) *Systematics and the Origin of Species* (Columbia Univ. Press, New York).
30. Simpson, G. G. (1944) *Tempo and Mode in Evolution* (Columbia Univ. Press, New York).
31. Orgel, L. E. (1963) *Proc. Natl. Acad. Sci. USA* **49,** 517–521.
32. Ninio, J. (1991) *Genetics* **129,** 957–962.
33. Molitor, M., Dahl, C., Molitor, I., Schafer, U., Speich, N., Huber, R., Deutzmann, R. & Trüper, H. G. (1998) *Microbiology* **144,** 529–541.
34. Langer, D., Hain, J., Thuriaux, P. & Zillig, W. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 5768–5772.
35. Olsen, G. J., Woese, C. R. & Overbeek, R. (1994) *J. Bacteriol.* **176,** 1–6.