RNA Biology 11:3, 1–6; March 2014; © 2014 Landes Bioscience

Looking in the right direction
Carl Woese and evolutionary biology

Nigel Goldenfeld

Institute for Universal Biology; Institute for Genomic Biology, and Department of Physics; University of Illinois at Urbana-Champaign; Urbana, IL USA

Carl Woese is known to the scientific community primarily through his landmark contributions to microbiology, in particular, his discovery of the third Domain of Life, which came to be known as the Archaea. While it is well known how he made this discovery, through the techniques he developed based on his studies of rRNA, the reasons why he was driven in this scientific direction, and what he saw as the principle outcome of his discovery—it was not the Archaea!—are not so widely appreciated. In this essay, I discuss his vision of evolution, one which transcends population genetics, and which has ramifications not only for our understanding of the origin of life on Earth and elsewhere, but also for our understanding of biology as a novel class of complex dynamical systems.

At exactly 2 pm on Friday September 20, 2002, I received the most important email of my life. Originating from a computer mysteriously called "ninja," the sender wasted no time on getting to the point:

"This is Carl Woese, over in Life Sciences. I'd like to talk to you at some point about moving the teaching of biology into the 21st century. Molecular biology clearly has lost (run out of) its vision, and a new and very different biology needs to emerge. I have been told of your interests, and know, thereofore [sic], that you are atune [sic] to what I'm talking about. I would like to see at least some cognizance on the part of card carrying biologists of complex dynamic systems, an appreciation for the fact that the cell is indeed a complex dynamic system and evolved in such a manner. My telephone is 3–9369, if you care to discuss the matter with me."

Naturally, I responded with alacrity, writing that

"I am a theoretical physicist, but tend to work on topics that are regarded as out of the mainstream by most of my colleagues … I don't know very much about biology, and worse, I don't think I have the sort of mind that can be engaged by or penetrate much of the subject. Despite these handicaps, …"

Carl's response was frank and, to be honest, tremendously exciting to me:

"You may not feel too much at home with biology as it now stands, but if I am any judge the field is decidedly moving to meet you."

Correspondence to: Nigel Goldenfeld; Email: nigel@uiuc.edu
http://dx.doi.org/10.4161/rna.28640
Submitted: 03/21/2014; Accepted: 03/21/2014

So began a scientific partnership and friendship that lasted more than a decade until his death. During that time, we met nearly every day and talked on the phone or via email otherwise. Looking back at these fragments of correspondence, it is remarkable to note how much of our future trajectory was set in those initial exchanges. Carl had indeed set his sights on a goal of making biology a quantitative science with roots in complex dynamical systems, but his enlisting a theoretical physicist to his cause was more than a way to help create a new breed of biologist—one with better math skills. Carl himself had trained as a physicist, with a BA in Mathematics and Physics from Amherst College in 1950 and a PhD in Biophysics from Yale three years later. Thus, he was no stranger to the great value that quantification could provide to biology. In fact, what Carl wanted was to complete his understanding of the evolutionary history of all life on Earth, a program of research that he had begun to think about seriously during the 1960s. That program of research had been articulated with clarity in a letter to Francis Crick dated June 24, 1969, a lengthy extract of which was reproduced in our article on the historical and conceptual relationship between microbiology, molecular biology, and evolution theory.[1]

"If we are ever to unravel the course of events leading to the evolution of the prokaryotic (i.e., simplest) cells, I feel it will be necessary to extend our knowledge of evolution backward in time by a billion years or so. i.e., backward into the period of actual Cellular Evolution."[1]

Carl had famously spent much of the following decade setting this investigation into motion, through his work on rRNA, culminating in the celebrated discovery of both the relatedness and tripartite structure of life by Woese and Fox in 1977,[2] and ultimately leading to a new proposal for the classification of life[3,4] that is today the mainstream view. Not so frequently emphasized, the finding that all life is related implies the existence of a last universal common ancestor (LUCA), now known to be positioned between the Bacterial and the Archaeal/Eukaryotic branches and representing in one extreme view a single organism, or in another view, a community of associated organisms. Although the significance of these two discoveries is hard to exaggerate, and despite Carl's manifest pride in this remarkable accomplishment, he was deeply dissatisfied with it, disappointed by a necessary limitation of his chosen instrument of biological revelation: the ribosome. The highly conserved nature of the ribosome, in particular 16SrRNA, made its molecular sequence a brilliant choice to mark the dynamics of evolution writ large; however, 16SrRNA sequence comparison only tracked (or more accurately, defined) the lineages of organisms whose cellular

structure included ribosomes. In the modern era that leaves out viruses of course, but in the era before LUCA, it necessarily leaves out all of life. LUCA was not just the last universal common ancestor but a representative of the first organisms which exhibited translational machinery in sufficiently advanced form to be traced by 16SrRNA phylogeny. Earlier classes of life would not have translational machinery recognizable as related to today's ribosomes, and thus would be invisible to 16SrRNA phylogeny. Woese's program to uncover the evolutionary history of life on Earth had apparently run into a roadblock.

The spectacular success of molecular phylogeny in uncovering the three Domains of Life overshadowed Carl's original endeavor. Carl was convinced that to get his original program on track, he would need to make a conceptual advance. One of my first questions to him was why are there only three Domains of Life? I was not attaching any special significance to the number three, but the fact that number is of order unity, and not one hundred, for example, is surely significant. What does this tell us about the singularity that is LUCA? What does it tell us about life before LUCA? Carl was extremely animated by these and related questions, but I soon found that there were more urgent preoccupations on his mind, because there were challenges to the very notion of Tree of Life itself, arising from the increasing recognition of the evolutionary impact of horizontal gene transfer (HGT). Remarkably, in understanding these issues, the life before LUCA problem also began to be resolved.

Horizontal gene transfer—the transmission of genes from one organism to another unrelated organism—was being invoked as an invalidation of the concept of lineage, thus casting doubt on the idea that post-LUCA, a Tree of Life was a meaningful concept (for a review, see ref. 5). Phylogenies based on genes other than those associated closely with translation and the ribosome clearly showed evidence for HGT, and non-canonical phylogenies, and this had led Carl and collaborators to dig into the history of the amino-acyl-tRNA-synthetases.[6] Carl had chosen the rRNA as likely to be the most conserved part of the translation machinery, and the synthetases represented the most likely component with any fluidity. The synthetase study showed that canonical pattern was certainly disrupted but the three Domains large-scale structure was still intact. Subsequent thoughtful analyses have basically confirmed this conclusion, although there remain many issues about which the community is still debating—such as whether or not HGT disrupts canonical pattern, or by virtue of it being more common among close relatives, actually enhances it (for summaries from a balanced perspective, see for example the paper and reviewer discussion in ref. 7 as well as related articles in refs. 8 and 9). Carl was not too impressed with the more extreme challenges to the three Domains of Life, such as those arising from whole genome phylogeny, because they showed a fundamental misunderstanding of what the concept of lineage meant: not all genes are equally appropriate for illuminating the large-scale structure of evolution. Nevertheless, he well understood the evolutionary power of HGT, and thus, was receptive to focusing not on the phylogeny arising from HGT but instead the rate of evolution.

Carl did not read Darwin's *Origin of Species* until around 2000, because what interested him more was not whether evolution had occurred, but the speed of evolution. In fact, the work that had made the most impression on him was G.G. Simpson's *Tempo and Mode of Evolution*,[10] because its analysis of the fossil record echoed something that had puzzled Carl right from the early results with Fox: How could evolution have achieved so much, starting from an abiotic earth and attaining an essentially modern translational machinery in a time frame of what could be at most one billion years? My early discussions with Carl centered on this issue, because Carl had the intuition that some understanding of complex dynamical systems would be pertinent to the issue. That is to say, he felt that the picture of evolution, which had emerged from the modern synthesis during the first half of the 20th century, was somehow missing an important aspect of the evolutionary process. One way to phrase this seeming inadequacy is to ask how population genetics can possibly be considered as a full explanation of the evolutionary process, when, by construction, the biological world before there were genes as such was manifestly beyond the regime of validity of the theory. Carl had already given a lot of thought to life before genes, and in the same year as the discovery of the Archaea had, in another magnificent paper also with Fox, initiated conceptual discussion on such a phase of life, which he called the "progenote."[11]

The progenote was a phase of life in which the distinction between genotype and phenotype had not yet emerged. Carl considered the emergence of the genotype–phenotype relationship to be the primary force shaping the evolution of the cell, an argument he based on the simple incisive observation that the translational apparatus is large and complex, perhaps more than any other cellular machinery. Carl attributed the size and complexity to the requirements for accuracy of the translational process, citing as inspiration a summary of theoretical work on cellular automata[12] conducted by John Von Neumann[13] during the phase of his life devoted to the construction of the world's first modern computer (for a stimulating account of this enterprise, see ref. 14). Elsewhere, I have described how Carl thought about this entity[15] in the context of his work on phylogeny, but here I want to focus on the issue of complexity and dynamics. Carl was sure that complex systems dynamics had something to do with the tempo and mode of evolution, but when pressed, he couldn't really say what this meant to him in detail or even why he had this feeling. Carl would often say that he lacked technical knowledge in biochemistry, structural biology, physics, and mathematics, but he more than made up for it with his imagination and intuition. This case was no exception, because our discussions quickly turned away from the vaguely defined notion of complexity, to the related but more specific question of collective effects.

Complex systems must be strongly interacting, so much so that the individual parts lose their identity and it is the relationship between those parts which is more important than the parts themselves. A typical paradigm is the idea of a network: nodes connected by links that constitute a model of a complex system such as metabolism or gene regulation. In the latter case, for example, the nodes are genes and the links might be transcription factors and gene expression levels of proteins. The nodes individually might have a complicated stochastic switching behavior, but the system as a whole can behave in some sort of synchrony to control the

cell cycle for example. The collective behavior of the whole system does not necessarily depend too much on the specific details of each node's dynamics and response to inputs, and this emergent phenomenon is one of the hallmarks of complex systems. Actually, Carl was so struck by this type of phenomenon that he eventually went out and bought half a dozen copies of Steven Strogatz's excellent book "Sync"[16] to give to friends and to distribute to students through the interdisciplinary library we set up at the Institute for Genomic Biology, where he worked for the past 7 y of his life.

Returning to the problem at hand, we realized that horizontal gene transfer also is a network effect, the nodes being the donor and recipient microbes, connected by a link that is the transferred gene. As recognized by earlier workers, the evolutionary impact of such a mechanism could be significant.[17-22] We began to wonder what emergent behavior would arise in populations whose evolution was dominated by HGT (a collective effect) as opposed to point mutation (a single body effect). In particular, we wondered how the translational machinery would evolve in the presence of HGT. HGT does not accomplish much unless there is a universal genetic code, so that the transferred gene can be expressed by the recipient organism. Could there be an autocatalytic, self-reinforcing mechanism by which an initial set of genetic codes, distributed across many competing cohorts of organisms, would themselves compete and lead to a universal genetic code? The more we discussed about these and other matters, it struck me how generic all these arguments were, not just ours, but all the "just so stories" about early evolution that were in the literature. Generic to me was a good thing, because it meant that if a mechanism could be formulated in the mathematics of dynamical systems, the outcomes could be enumerated by simulation or other analysis, and the robustness checked. Here at last was a way that theoretical physics could perhaps make a contribution!

The task of turning these early ideas into a concrete research project was not trivial, but this was accomplished with the participation of Kalin Vetsigian, at that time a brilliant graduate student working with me, who has gone on to an appropriately distinguished career at Harvard and Wisconsin. Central to our work was the idea that working with minimal models of the evolution of translation meant that we needed to have a control model (something like a theorist's version of a null hypothesis), so that by comparing two minimal models with different ingredients included in the dynamics, we could isolate the dynamical contributions of the ingredients separately. For us, the control model was very clear: evolution of translation with purely vertical "Darwinian evolution." That is, we would explore the dynamics of early organisms, competing and trying to occupy niche environments, in two ways: (1) adapting by mutation alone, or (2) with the added ingredient of HGT, but an HGT whose efficacy would actually emerge dynamically from the calculation and would not be put in by hand as it were. To our astonishment, a very simple conclusion quickly became apparent. Emergent HGT accelerated the dynamics of evolution, because of the collective network effect. This was not a surprise, because we had already anticipated that this could arise. What was a surprise however was that the resulting translational machinery, as represented by the abstraction of a genetic code, evolved in specificity and precision, leading to a code that was in some sense optimal (or very nearly so) in mitigating the effects of errors in translation and indeed point mutations of the genome. This "optimality" had been guessed by Woese in a remarkable paper,[23] which also was the first to describe a dynamical scenario of the gradual refinement of a translational machinery from being able to produce statistical distributions of proteins to the present day highly deterministic machinery necessary for complex life. The idea that the modern canonical code itself would be nearly optimal had subsequently been rediscovered and quantified by Haig and Hurst, and later others, using Monte Carlo sampling of synthetic genetic codes that differed from the canonical one through permutation of the amino acids in the code table (although retaining the canonical degeneracy structure).[24-30]

That this arose from HGT but not vertical evolution was to us quite remarkable, because it (and later refinements to take into account tRNA abundance[31]) solved in one stroke the key facets of the evolution of the genetic code. First of all, our model and the preceding one of Ardell and Sella[32-34] were concrete counterarguments to Crick's dismissal of an evolved code through his memorable term "frozen accident." Carl's relationship with Crick was an interesting and complex one, and this is not the place to go into it in detail. However, Carl got a kick out of showing that Crick's argument was wrong, and the reasons for this only became clear to me recently. Previously, I referred to Carl's letter to Crick, in which he announced his intention to use molecular sequences to map out the evolutionary history of life. Crick's response was not fully supportive, to my surprise; he expressed doubts about the evolutionary program, and suggested that it be hedged with a study of sequence and function, something that I know now to have been anathema to Carl. Crick wrote:

"I think the project is a good one and an important one but it may well be difficult to get money for it, especially as its rather a gamble whether enough evidence is still frozen in the sequences. For this reason, I suggest you draw up a slightly enlarged programme—to study the effects of change of sequence on function, which is bound to give results even if the evolutionary results prove disappointing. Such a combined project might well attract enough money to finance it."— F. Crick unpublished letter (1969).

Second, our work finally answered Carl's overriding preoccupation with the tempo and mode of evolution: yes, both the mode and tempo were different before LUCA, with a rapidly evolving collective state generating genetic novelty in an exponential growth process that culminated through mechanisms still not properly understood in the transition to vertical evolution and the emergence of three Domains of Life at LUCA. Third, LUCA was identified to be not a single organism but the residue of a collective state of life, one with rampant HGT: Carl's mysterious progenote, finally interpreted dynamically. Fourth, our work showed that genetic code would be both universal and nearly optimal, not due to fine tuning or chemical properties of molecules, but because these characteristics were the dynamical attractors of a co-evolutionary process of code refinement and organismal complexification. This process is something that is generic to what we termed innovation-sharing protocols, and is an example of what has been presciently called "universal biology."[35]

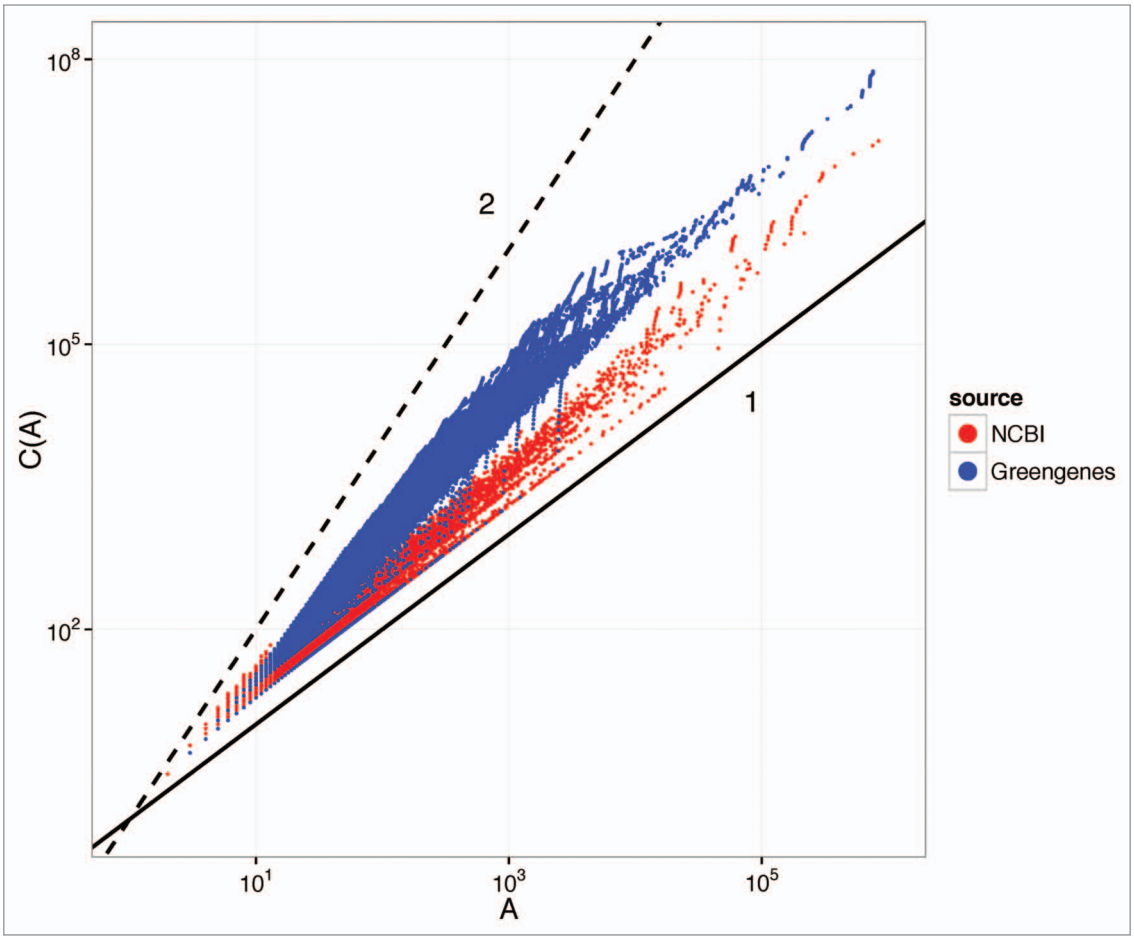

Figure 1. Scaling properties of phylogenetic and taxonomic trees, as described in the text. The cumulative subtree size is plotted as a function of the subtree size for each node of a tree. Red, trees constructed from the NCBI database. Blue, trees constructed from the Greengenes database. The figure shows that taxonomic trees are more balanced than evolutionary trajectories as measured by 16SrRNA sequences.

It is no exaggeration to say that Carl was very thrilled by these results. Although I feel that the roots of this work were long anticipated by Carl, the remarkable inevitability of the dynamical mechanism for code evolution was not something that had been forseen. I believe that the power of reasoning using mathematics and dynamical systems theory was something that Carl found truly fulfilling, in some sense a vindication of his long journey as a physicist wondering within the world of biology.

Carl's lonely journey is well-known of course,[36] but perhaps less well-appreciated is his enduring self-identification as a physicist, something that became a scientific issue in his famous skirmish with Ernst Mayr—a true battle of titans undertaken in the pages of Nature and the Proceedings of the National Academy of Sciences.[37-39] Paradoxically, Mayr had been Carl's strong supporter and indeed had been the one who nominated Carl to the National Academy of Sciences. However, as the results of Carl's program became apparent, Mayr recoiled from the microbe-centric view of life that emerged. He challenged Carl's three Domains, and the concomitant refutation of the concept of the prokaryote–eukaryote distinction as being the only meaningful one in biological classification. In order to buttress his arguments, he pointed out that Carl's heritage as a physicist had influenced his choice of molecular techniques for tracing the course of evolution and classifying lineages by their evolutionary trajectory. If Carl had been a biologist, with an appreciation of the importance of organismal features identified through the practice of taxonomy, he would not have made the mistake of proposing the three Domains classification of life. Carl's rebuttal is a beautiful scientific argument, and of course the debate ended with Mayr's passing. However, I had the privilege of making a contribution to this conversation, and want to end with another anecdote about that, one that we might call Woese's last blast.

For various reasons, Carl and I had become interested in the new science of metagenomics as it developed during the late 2000s. I became curious about the statistical properties of phylogenetic trees, and as genomes accumulated, it became possible to look for scaling laws in phylogeny, in order to get some understanding of the statistical dynamics of the evolutionary process itself. In order to do this, I started thinking about the topological properties of branching phylogenetic trees and came up with some ways to quantify this, which were simultaneously proposed by Herrada et al.[40] The basic idea is to measure two quantities on each node of a phylogenetic tree. The first we will call A: the subtree size (ie. the number of subtaxa diversifying from the node). The second

we will call the cumulative branch size C: the sum of the branch sizes of all the subtree nodes. This quantity measures the shape of the subtree in some sense. For symmetric, equally balanced trees, C is the smallest possible value, whereas for the most asymmetric, comb-like subtree, with branches splitting all the time from a single branch, C takes on its largest possible value. The question is: are C and A related? Herrada et al.[40] found that C and A were essentially related by a power law relationship: log C/log A ~z where the exponent z can be between 1 and 2. In practice z ~1.4, a result confirmed and extended by another outstanding student, Patricio Jeraldo, who has explored this question in great detail in his unpublished PhD thesis.[41] It is interesting to ask what happens if trees are constructed from the NCBI taxonomic database, rather than the set of fully sequenced genomes. Patricio discovered that the result also is a power law, but the exponent z is very close to unity. In other words, the topological structure of phylogenetic trees and taxonomic trees are very different, and this is reflected in their statistical scaling properties, as shown in **Figure 1**. The result was especially puzzling because the value of the exponent z being close to unity suggested that the taxonomic trees were somehow more balanced than phylogenetic trees. Why? As soon as Patricio's analysis was plotted, I called up Carl, and asked him to come into the lab to see something interesting. He arrived shortly and I explained the calculation and the interpretation for him, drawing out a caricature of what the trees would look like corresponding to different values of the exponent z. His reaction was immediate and fascinating: "Ah," he said, "I think you've discovered Mayr's Principle of Balance!" Mayr's Principle of Balance[38] states that "the retrieval of information is greatly facilitated if the taxa at a given categorical rank are, as far as possible, of equal size and degree of diversity," a principle that arises from Mayr's view that the purpose of classification is to facilitate information retrieval.[42] Carl's view, of course, was that of an evolutionist, not a collector: the purpose of classification is to group organisms together that shared a common evolutionary lineage. **Figure 1** dramatically captures the tension between these two viewpoints: the fact that taxonomy exhibited a scaling law with exponent z close to unity is an artifact of Mayr's approach to classification, reflecting the Principle of Balance that has apparently been put in by hand into taxonomic classification. The fact that the phylogenetic trees exhibit a different scaling exponent close to 1.4 indicates the natural dynamics of the evolutionary processes, and reflects the vast difference between Carl's evolutionary perspective on biology and Mayr's classification-based perspective.

Although Carl considered this analysis as some sort of vindication of his views—nature's classification vs. human classification—he did not gloat. Carl had vanquished the ghosts of both Crick and Mayr, but what was important to him was the scientific understanding that emerged. That's who he was. A true scientist.

When I think of Carl Woese, the scientist, he is in his office at Morrill Hall during those early sessions in September 2002, lounging indecorously in his ancient chair, the one with a thick pad of washroom towels taped to the arms as a primitive and zero-budget forerunner of ergonomic design. His feet are up on the desk, really a lab bench. Later he is pecking at a grubby keyboard attached to a SUN workstation—the famed "ninja" no less—that is a decade's worth of Moore's Law behind the times. Behind him is a giant poster of Miles Davis and an American flag adorns the wall facing his lab bench. In this room, 25 y previously, Carl unraveled the history of life on Earth. Or at least the last 3.8 billion years of it. This is to be the day that I receive my first biology lesson. I make a careful but genuine expression of respect and gratitude, to let him know that I know who he is. He acknowledges with a slight lowering of the head, the eyes closing gently, and then he starts to talk. He has no interest in telling me what I can read in books and papers. He wants to know what I think. Clearly this is a man without pretentions.

When I think of my friend Carl Woese, he is standing in the hallway between our offices in the IGB. He is laughing loudly and raucously at a joke, a sound like a seal barking, and beating the wall with his hand. Carl saw no reason not to have fun while we worked. He had perhaps the cleverest and most idiosyncratic sense of humor that I have ever encountered: subdued, subtle, and witty or off-scale rambunctious and sometimes even bawdy. Very little in between. He didn't tell jokes. He jested. Carl had a sub-clinical Asperger's type of personality, and this meant that he was unable to put himself in the mind of the other person. His humor was, thus, sometimes unfettered by good taste; but it always betrayed the fact that he looked at life in a very different way than did other people.

When I think of Carl Woese, my collaborator, we are sitting at my desk in the physics department going over the final version of our genetic code paper. It has already taken us a year to write. We know that it's going to be a good paper, perhaps even a very good one. But right now, we are arguing over commas. If you ever heard Carl give a lecture or introduce a visitor, you'll know that in public he was not an articulate speaker. But, man, could he write! And no detail was too small to be gone over and again and honed until it was perfect—a feedback loop that sometimes seemed never-ending. Carl's ability to focus was simply extraordinary, and of course it is that which got him through the 10 y in the wilderness, performing mind-numbingly dull experiments to create the molecular catalogs from which he deduced the history of life. In his personal reminiscences, he recalls saying to himself as he went home in the evening: "Woese, you have destroyed your mind again today."

When I think of Carl Woese, the man, he is standing in the kitchen of his house talking to my daughter Zippy. She's a horse-rider, and so it turns out, was Carl once. His face wears a beatific smile and his eyes are far away. I never saw him so happy as that moment when he was remembering riding horses in his youth. That was Carl's "Rosebud" moment.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

It is a pleasure to acknowledge the collaboration of Patricio Jeraldo in the work briefly summarized in **Figure 1** and described in the text. This material is based upon work supported by the National Aeronautics and Space Administration through the NASA Astrobiology Institute under Cooperative Agreement No. NNA13AA91A issued through the Science Mission Directorate.

References

1. Woese CR, Goldenfeld N. How the microbial world saved evolution from the scylla of molecular biology and the charybdis of the modern synthesis. Microbiol Mol Biol Rev 2009; 73:14-21; PMID:19258530; http://dx.doi.org/10.1128/MMBR.00002-09
2. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A 1977; 74:5088-90; PMID:270744; http://dx.doi.org/10.1073/pnas.74.11.5088
3. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 1990; 87:4576-9; PMID:2112744; http://dx.doi.org/10.1073/pnas.87.12.4576
4. Wheelis ML, Kandler O, Woese CR. On the nature of global classification. Proc Natl Acad Sci U S A 1992; 89:2930-4; PMID:11537862; http://dx.doi.org/10.1073/pnas.89.7.2930
5. Gogarten JP, Doolittle WF, Lawrence JG. Prokaryotic evolution in light of gene transfer. Mol Biol Evol 2002; 19:2226-38; PMID:12446813; http://dx.doi.org/10.1093/oxfordjournals.molbev.a004046
6. Woese CR, Olsen GJ, Ibba M, Söll D. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol Mol Biol Rev 2000; 64:202-36; PMID:10704480; http://dx.doi.org/10.1128/MMBR.64.1.202-236.2000
7. Andam CP, Gogarten JP. Biased gene transfer and its implications for the concept of lineage. Biol Direct 2011; 6:47; PMID:21943000; http://dx.doi.org/10.1186/1745-6150-6-47
8. Andam CP, Williams D, Gogarten JP. Biased gene transfer mimics patterns created through shared ancestry. Proc Natl Acad Sci U S A 2010; 107:10679-84; PMID:20495090; http://dx.doi.org/10.1073/pnas.1001418107
9. Andam CP, Williams D, Gogarten JP. Natural taxonomy in light of horizontal gene transfer. Biol Philos 2010; 25:589-602; http://dx.doi.org/10.1007/s10539-010-9212-8
10. Simpson GG. Tempo and mode in evolution. Columbia University Press, 1944.
11. Woese CR, Fox GE. The concept of cellular evolution. J Mol Evol 1977; 10:1-6; PMID:903983; http://dx.doi.org/10.1007/BF01796132
12. Burks AW. Von Neuman's Self-Reproducing Automata. Urbana, IL: University of Illinois Press, 1970.
13. Von Neumann J. Theory of Self-Reproducing Automata. Urbana, IL: University of Illinois Press, 1966.
14. Dyson G. Turing's Cathedral: The Origins of the Digital Universe. Random House LLC, 2012.
15. Pace NR, Sapp J, Goldenfeld N. Phylogeny and beyond: Scientific, historical, and conceptual significance of the first tree of life. Proc Natl Acad Sci U S A 2012; 109:1011-8; PMID:22308526; http://dx.doi.org/10.1073/pnas.1109716109
16. Strogatz S. Sync: The emerging science of spontaneous order. Hyperion, 2003.
17. Anderson ES. Possible importance of transfer factors in bacterial evolution. Nature 1966; 209:637-8; PMID:5921205; http://dx.doi.org/10.1038/209637a0
18. Sonea S. A bacterial way of life. Nature 1988; 331:216; PMID:3336435; http://dx.doi.org/10.1038/331216a0
19. Syvanen M. Cross-species gene transfer; implications for a new theory of evolution. J Theor Biol 1985; 112:333-43; PMID:2984477; http://dx.doi.org/10.1016/S0022-5193(85)80291-5
20. Syvanen M. Horizontal gene transfer: evidence and possible consequences. Annu Rev Genet 1994; 28:237-61; PMID:7893125; http://dx.doi.org/10.1146/annurev.ge.28.120194.001321
21. Syvanen M. Evolutionary implications of horizontal gene transfer. Annu Rev Genet 2012; 46:341-58; PMID:22934638; http://dx.doi.org/10.1146/annurev-genet-110711-155529
22. Anderson NG. Evolutionary significance of virus infection. Nature 1970; 227:1346-7; PMID:5455138; http://dx.doi.org/10.1038/2271346a0
23. Woese CR. On the evolution of the genetic code. Proc Natl Acad Sci U S A 1965; 54:1546-52; PMID:5218910; http://dx.doi.org/10.1073/pnas.54.6.1546
24. Haig D, Hurst LD. A quantitative measure of error minimization in the genetic code. J Mol Evol 1991; 33:412-7; PMID:1960738; http://dx.doi.org/10.1007/BF02103132
25. Freeland SJ, Hurst LD. The genetic code is one in a million. J Mol Evol 1998; 47:238-48; PMID:9732450; http://dx.doi.org/10.1007/PL00006381
26. Butler T, Goldenfeld N, Mathew D, Luthey-Schulten Z. Extreme genetic code optimality from a molecular dynamics calculation of amino acid polar requirement. Phys Rev E Stat Nonlin Soft Matter Phys 2009; 79:060901; PMID:19658466; http://dx.doi.org/10.1103/PhysRevE.79.060901
27. Freeland SJ, Knight RD, Landweber LF, Hurst LD. Early fixation of an optimal genetic code. Mol Biol Evol 2000; 17:511-8; PMID:10742043; http://dx.doi.org/10.1093/oxfordjournals.molbev.a026331
28. Knight RD, Freeland SJ, Landweber LF. Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet 2001; 2:49-58; PMID:11253070; http://dx.doi.org/10.1038/35047500
29. Itzkovitz S, Alon U. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Res 2007; 17:405-12; PMID:17293451; http://dx.doi.org/10.1101/gr.5987307
30. Goldman N. Further results on error minimization in the genetic code. J Mol Evol 1993; 37:662-4; PMID:8114119; http://dx.doi.org/10.1007/BF00182752
31. Vetsigian K, Goldenfeld N. Genome rhetoric and the emergence of compositional bias. Proc Natl Acad Sci U S A 2009; 106:215-20; PMID:19116280; http://dx.doi.org/10.1073/pnas.0810122106
32. Ardell DH, Sella G. On the evolution of redundancy in genetic codes. J Mol Evol 2001; 53:269-81; PMID:11675587; http://dx.doi.org/10.1007/s002390010217
33. Ardell DH, Sella G. No accident: genetic codes freeze in error-correcting patterns of the standard genetic code. Philos Trans R Soc Lond B Biol Sci 2002; 357:1625-42; PMID:12495519; http://dx.doi.org/10.1098/rstb.2002.1071
34. Sella G, Ardell DH. The coevolution of genes and genetic codes: Crick's frozen accident revisited. J Mol Evol 2006; 63:297-313; PMID:16838217; http://dx.doi.org/10.1007/s00239-004-0176-7
35. Des Marais DJ, Nuth JA 3rd, Allamandola LJ, Boss AP, Farmer JD, Hoehler TM, Jakosky BM, Meadows VS, Pohorille A, Runnegar B, et al. The NASA astrobiology roadmap. Astrobiology 2008; 8:715-30; PMID:18793098; http://dx.doi.org/10.1089/ast.2008.0819
36. Morell V. Microbiology's scarred revolutionary. Science 1997; 276:699-702; PMID:9157549; http://dx.doi.org/10.1126/science.276.5313.699
37. Mayr E. A natural system of organisms. Nature 1990; 348:491; PMID:1701032; http://dx.doi.org/10.1038/348491a0
38. Mayr E. Two empires or three? Proc Natl Acad Sci U S A 1998; 95:9720-3; PMID:9707542; http://dx.doi.org/10.1073/pnas.95.17.9720
39. Woese CR. Default taxonomy: Ernst Mayr's view of the microbial world. Proc Natl Acad Sci U S A 1998; 95:11043-6; PMID:9736686; http://dx.doi.org/10.1073/pnas.95.19.11043
40. Herrada EA, Tessone CJ, Klemm K, Eguíluz VM, Hernández-García E, Duarte CM. Universal scaling in the branching of the tree of life. PLoS One 2008; 3:e2757; PMID:18648500; http://dx.doi.org/10.1371/journal.pone.0002757
41. Jeraldo P. Computational approaches to stochastic systems in physics and ecology. Department of Physics. Urbana, IL: University of Illinois at Urbana-Champaign, 2012.
42. Mayr E. Systems of ordering data. Biol Philos 1995; 10:419-34; http://dx.doi.org/10.1007/BF00857592