REVIEW: EVOLUTION

Planetary Biology—Paleontological, Geological, and Molecular Histories of Life

Steven A. Benner,[1,2,3,4*] M. Daniel Caraco,[1,4,5] J. Michael Thomson,[2] Eric A. Gaucher[3]

The history of life on Earth is chronicled in the geological strata, the fossil record, and the genomes of contemporary organisms. When examined together, these records help identify metabolic and regulatory pathways, annotate protein sequences, and identify animal models to develop new drugs, among other features of scientific and biomedical interest. Together, planetary analysis of genome and proteome databases is providing an enhanced understanding of how life interacts with the biosphere and adapts to global change.

A key goal for biology in the postgenomic era is to use the sequences of genes and proteins to generate information about molecular, cellular, and organismal biology. In the future as in the past, much of this information will undoubtedly be obtained through biochemical, genetic, and molecular biological experiments in the laboratory. This notwithstanding, almost any approach that provides inferences, insights, or information about biology from sequence data without requiring additional experimentation will be valuable.

For this reason, considerable attention has been directed toward the fact that biomolecular sequences contain information about their historical past (*1*). The search for homologs, or protein sequences that diverged from a common ancestor, is frequently the principal tool used to annotate sequence databases. Likewise, the sequences of a set of homologous proteins suggest a tree that defines the history of the protein family. These trees can be used to infer familial relationships between the organisms that carry the proteins, define the order in which particular taxa diverged (*2, 3*), constrain the connectivity of the deep branches joining the primary kingdoms of life (*4*), and even correlate the divergence of species with their migrations across drifting continents (*5*).

This theme has been amplified by recognizing that two other fields, geology and paleontology, also provide records of the history of life on Earth. In many respects, these records complement the record contained in molecular sequence data. For this reason, considerable effort is now been directed toward explicit connection of these three records. Here, the past is the key to the present. By understanding the history by which a protein emerged within the context of its planetary biology, we hope to better understand how it functions in contemporary life.

Joining Records Through Dating

It is not easy, of course, to connect records that lie within rocks and bones with a record that is captured in the sequences of organic molecules, proteins, and the DNA that encodes them. One way to do so is to use historical dates as the connector. Radiochemical dating, used to date events in the geological record, offers the gold standard. Radioisotopes decay via a first-order process. The amount of isotope remaining after time t follows a simple exponential rate law, with the fraction of initial atoms remaining $f = 1 - \exp(-kt)$, where k is the rate constant for the decay process, and the half-life $\tau = \ln 2/k$. Using two isotopes of uranium in a zircon from an igneous rock, for example, precision to better than a million years is routine for a rock 500 million years (Ma) old (*6*).

This precision is envied by those who date events in the paleontological and molecular records. In paleontology, the rate of morphological change cannot follow exponential kinetics, as it does not approach an end point. Rocks that contain fossils are occasionally associated with rocks that carry datable radioisotopes, of course. These permit accurate limits to be set for dates for specific fossils in specific strata. Unfortunately, the fossil record is incomplete, meaning that we rarely (if ever) date fossils that represent branch points in a phylogenetic tree, or the first appearance of a species. This incompleteness creates large uncertainties in dates at nodes in trees, even if we have good dates for the rocks that contain the fossils.

Dating events in the molecular record is still more problematic. In the 1970s, many groups explored the possibility of constructing a molecular clock by counting replacements separating two sequences and assuming that the rate constant for amino acid replacement is invariant over time (*7*). Unfortunately, protein behaviors are too closely tied to the demands and constraints of natural selection. Amino acid replacement is faster or slower depending on how these change, making protein sequences irregular clocks at best (*8*). Collections of protein sequence families might be used to date the divergence of taxa, in the hope that this episodic rate variation averages over the collection to give an apparently time-invariant rate constant (*9, 10*). But individual protein sequences generally serve as poor clocks, and it is difficult to

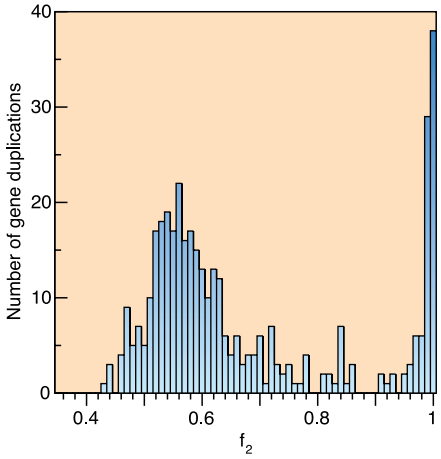

Fig. 1. The number of duplicated gene pairs (vertical axis) in the genome of the yeast *Saccharomyces cerevisiae* versus f_2, a metric that models divergence of silent positions in twofold redundant codon systems via an approach-to-equilibrium kinetic process and therefore acts as a logarithmic scale of the time since the duplications occurred. Recent duplications are represented by bars at the right. Duplications that diverged so long ago that equilibrium at the silent sites has been reached are represented by bars where $f_2 \approx 0.55$. Noticeable are episodes of gene duplication between the two extremes, including a duplication at $f_2 \approx 0.84$. This represents the duplication, at ~80 Ma, whereby yeast gained its ability to ferment sugars found in fruits created by angiosperms. Also noticeable are recent duplications of genes that enable yeast to speed DNA synthesis, protein synthesis, and malt degradation, presumably representing yeast's recent interaction with humans.

[1]Department of Chemistry, [2]Department of Anatomy and Cell Biology, and [3]NASA Astrobiology Institute, Department of Chemistry, University of Florida, Gainesville FL, 32611–7200, USA. [4]Foundation for Applied Molecular Evolution, Post Office Box 13174, Gainesville FL, 32604, USA. [5]Department of Chemistry, ETH Hönggerberg, CH-8093 Zürich, Switzerland.

*To whom correspondence should be addressed. E-mail: benner@chem.ufl.edu

correlate the molecular record for a specific protein family with the paleontological record.

To minimize the effects of selective pressure as they construct molecular clocks, most workers now examine silent sites in a gene (*11*). Because they do not change the coding sequence of a protein, nucleotide substitutions at silent sites cannot alter the behavior of a protein. They are therefore most likely to be free of the selective pressures that cause protein clocks to "tick" irregularly.

Silent sites are of many types in the standard genetic code. Some offer better clocks than others. Most useful are silent sites in codon systems that are twofold redundant (*12*). Here, exactly two codons encode the same amino acid. These codons are interconverted by transitions: a pyrimidine replacing another pyrimidine, or a purine replacing another purine. When the amino acid itself is conserved, the divergence at such sites can be modeled as an "approach to equilibrium" kinetic process, just as radioactive decay, with the end point being the codon bias b. Here, the fraction of paired codons that are conserved $f_2 = b + (1 - b)\exp(-kt)$, where again k is the first-order rate constant and t is the time. Given an estimate of the rate constant k for these "transition-redundant approach-to-equilibrium" processes, if k and b are time-invariant, one can estimate the time t for divergences of the two sequences. An empirical analysis suggests that codon biases and rate constants for transitions has been remarkably stable, at least in vertebrates, for hundreds of millions of years (*12*). Therefore, approach-to-equilibrium metrics provide dates for events in molecular records within phyla, especially of higher organisms. These dates are useful to time-correlate events in the molecular record with events in the paleontological and geological records.

Identifying Pathways from Genomic Records

Simultaneous events need not, of course, be causally related, especially when simultaneity is judged using dating measurements with variances of millions of years. But observation that two events in the molecular record are nearly contemporaneous suggests, as a hypothesis, that they might be causally related. Such hypotheses are testable, often by experiment, and are useful because they focus experimental work on a subset of what would otherwise be an extremely large set of testable hypotheses.

Consider, for example, the yeast *Saccharomyces cerevisiae*, whose genome encodes ~6000 proteins. The yeast proteome has 36 million potentially interacting pairs. Some investigators in the field of systems biology are laboring to experimentally examine all of these in the laboratory, hoping to identify these interactions (*13, 14*).

Correlating dated events in the molecular record offers a complementary approach. Gene duplications generate paralogs, which are homologous proteins within a single genome. Paralogous sequences can be aligned, their f_2 calculated, and their divergence dated. In yeast, paralog generation has occurred throughout the historical past (Fig. 1). A prominent episode of gene duplication, however, is found with an f_2 near 0.84, corresponding to duplication events that occurred ~80 Ma, based on clock estimates that generated divergence dates in fungi (*15*).

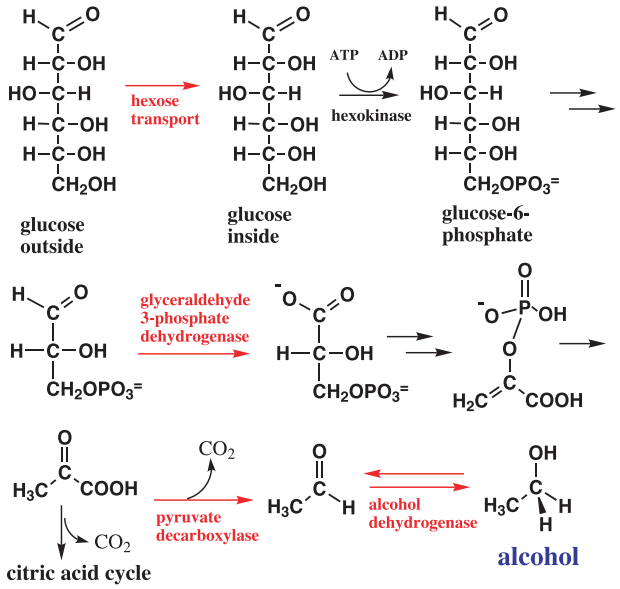

Fig. 2. The chemical pathway that converts glucose to alcohol in yeast arose ~80 Ma, near the time that fermentable fruits became dominant. Gene families that suffered duplication near this time, captured in the episode of gene duplication represented in the histogram in Fig. 1 by bars at $f_2 \approx 0.84$, are named in red. According to the hypothesis, this pathway became useful to yeast when angiosperms (flowering, fruiting plants) began to provide abundant sources of fermentable sugar in their fruits. Here, planetary biology has given meaning to a pathway that, without it, is memorized by generations of biochemistry students.

These duplications created several new sugar transporters, new glyceraldehyde-3-phosphate dehydrogenases, the nonoxidative pyruvate decarboxylase that generates acetaldehyde from pyruvate, a transporter for the thiamine vitamin that is used by this enzyme, and two alcohol dehydrogenases that interconvert acetaldehyde and alcohol.

This is not a random collection of proteins. Rather, these proteins all belong to the pathway that yeast uses to ferment glucose and produce ethanol (Fig. 2). Correlating the times of duplication of genes in the yeast genome has identified a pathway.

The approach-to-equilibrium dating tools can be more effective at inferring possible pathways from sequence data than can other approaches (*16–18*), especially for recently evolved pathways. By adding the geological and paleontological records to the analysis, however, these pathways assume additional biological meaning. Fossils suggest that fermentable fruits also became prominent ~80 Ma, in the Cretaceous, during the age of the dinosaurs (*19*). Indeed, overgrazing by dinosaurs may explain why flowering plants flourished (*20, 21*). Other genomes evidently also record episodes of duplication near this time, including those of angiosperms (which create the fruit) and fruit flies (whose larvae eat the yeast growing in fermenting fruit) (*22, 23*).

Thus, time-correlation among the three records connected by approach-to-equilibrium dates generates a planetary hypothesis about function of individual proteins in yeast, one that goes beyond a statement about a behavior ("this protein oxidizes alcohol . . . ") and a pathway (" . . . acting with pyruvate decarboxylase . . . ") to a statement about planetary function (" . . . allowing yeast to exploit a resource, fruits, that became available ~80 Ma"). This level of sophistication in the annotation of a gene sequence is difficult to create in any other way.

Approach-to-equilibrium dating methods are limited by the fact that silent clocks "tick" fast, meaning that information about ancestral species is rapidly lost. In vertebrates, a typical single lineage rate constant for silent substitution is 3×10^{-9} changes/site/year. This corresponds to a half-life of about 260 million years, which provides a time scale where this tool is optimal for dating. Various approaches are conceivable to extend dates obtained from these methods back to perhaps 500 Ma in vertebrates.

Such dates prove to be interesting for biomedical research. Virtually all function peculiar to vertebrates and their associated diseases arose in the past 500 million years, including cardiovascular disease, inflammation, autoimmune diseases, pain and other neurological disorders, and certain cancers. In each, targets must be identified, animal models chosen, research directions tested, function assigned, and pathways explicated. Correlating the three records generates understanding of physiology, function, and disease that directs this effort.

Resurrecting Ancient Proteins from Extinct Organisms

The emergence of angiosperms is only one of many changes in the history of the bio-

Fig. 3. *Leptomeryx* from the ~35 Ma Oligocene (Nebraska, courtesy S. A. Benner collection) was a mammalian ruminant herbivore representing approximately an ancestor of deer, giraffe, antelope, and ox. By hypothesis, the ruminant digestion of *Leptomeryx* offered it advantages in digesting tough grasses, which gained prominence through the Oligocene global cooling. The flowering, fruiting angiosperms behind the fossil, now presenting berries, are descendants of plants that provided the resource for the emergence of fermentation in yeast (Fig. 2) ~50 million years earlier.

sphere that left a record in the genomes of the surviving organisms. Consider, for example, the Oligocene epoch, which began ~35 Ma. The dinosaurs were extinct by the start of the Oligocene, mammals had come to occupy the ecological niches that the dinosaurs had previously enjoyed, and Earth was largely a tropical rain forest (*24*). During the Oligocene, however, the planet began to cool; its mean temperature dropped by perhaps 15°C (*25*). Various explanations for the cooling have been proposed, including changes in the position of continents and seaways that separate them, changes in atmospheric carbon dioxide concentrations, cosmological events (supernovae or impacts), and changes in the planet's orbit (*26*).

For whatever reason, the succulent vegetation of the rain forest was replaced by low-nutrition, silica-containing grasses over much of the planet. The large herbivorous mammals responded. Nonruminant herbivores (the brontotheres, hyracodons, and others) came to be displaced by ruminant mammals (camels, deer, and bovids), perhaps because of the efficiency with which ruminants digest grass (Fig. 3) (*27*). Ruminants have a first stomach that holds bacteria that ferment grass. The bacteria from the first stomach are then "eaten" in the second stomach, where their cell walls are broken by lysozymes, their nucleic acids are degraded by ribonucleases, and their proteins are digested by proteases.

Each of these protein families suffered duplications and/or rapid evolution near the time that ruminant digestion originated. To analyze these events, the sequences of ancestral proteins at nodes in their evolutionary trees are reconstructed, by inference from descendant sequences and using models for how protein sequences diverge in general (Fig. 4) (*28*). This process is much like the process used by historical linguists when reconstructing the Proto–Indo-European language from its descendant languages (*29*).

Explicit reconstructions of evolutionary intermediates assign specific amino acid replacements to specific episodes in the history of a protein family. In ancestral lysozyme genes, for example, rapid sequence evolution occurred as ruminant and ruminant-like digestion emerged (*30*). Rapid change in the sequence of a lysozyme implies rapid change in the behavior of lysozyme, which in turn suggests a change in its functional behavior. This hypothesis is inferential, of course, but can be tested. Further, it makes sense in light of a historical model. New lysozymes are expected to emerge to break open bacterial cells in the new ruminant digestion.

One way of testing such hypotheses is to resurrect the ancestral proteins and study their behavior in the laboratory. To do this, a DNA molecule encoding the ancestral protein is synthesized and expressed in an appropriate host. The ancient protein is then recovered and studied to determine whether its properties are consistent with its inferred ancestral role.

A paleobiochemical experiment was done for ruminant digestive ribonucleases (*31*), which also suffered gene duplication and an episode of rapid sequence evolution in the Oligocene. Laboratory studies on ancient ribonucleases found that the substrate specificity and stability of the emerging ruminant ribonuclease changed in this episode in a way consistent for a protein being recruited to play a new role in the digestive tract. These data added a planetary dimension to the annotation of the ribonuclease protein. Rather than saying that "ribonuclease is involved in ruminant digestion," we can say that digestive ribonuclease emerged near the time when ruminant digestion emerged, in animals in which ruminant digestion developed, at a time where difficult-to-digest grasses emerged, permitting their descendants to exploit a newly available resource emerging at a time of global climatic upheaval.

Functional inference from reconstructed evolutionary biology (*32*) has been applied to aromatases in swine (the three paralogs appear to manage large litter sizes) (*33*), bacterial dehydrogenases (guiding protein engineering experiments) (*34*), elongation factors (showing a change in functional behavior in a family that was widely regarded as not having undergone adaptive evolution) (*35*), mammalian kinases, mammalian SH2 domains, and mammalian cytokines (showing the connection between structural biology and adaptive recruitment) (*32*), as well as snake venom phospholipases (correlating functional changes with evolution of prey) (*36*). In many of these examples, more reliable dating generated better correlation among the molecular, paleontological, and geological records. This correlation, in turn, suggested chemical experiments to better capture the underlying reactivity of the proteins, assign roles to specific residues involved in the changing function, and test hypotheses relating structure and behavior to changing function.

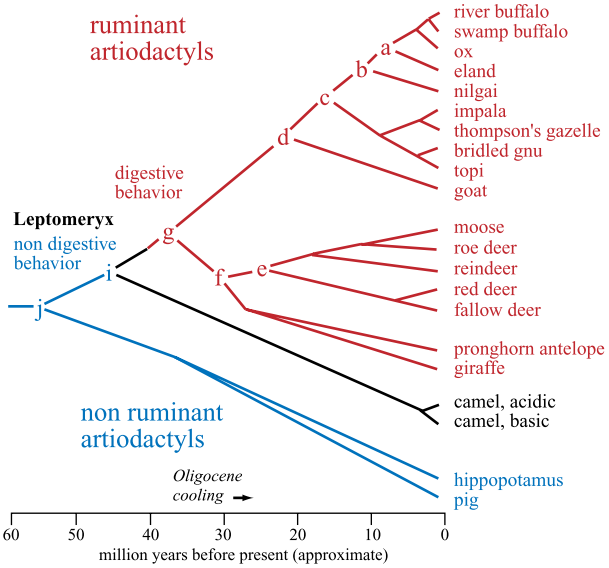

Fig. 4. An evolutionary tree relating ribonucleases responsible for the digestion of nucleic acid from bacteria fermenting grass in the first stomach of ruminants. Experimental analysis of reconstructed ancestral proteins suggests that digestive-like behavior in the protein arose near the time that its putative role in digestion in ruminants arose, near the time when grasses arose in response to the global climatic upheaval known as the Oligocene cooling.

Pharmacological Models and Biomedical Research

A historical approach to contemporary biology is relevant to biomedical research, in part because it helps define the function of proteins. In general, proteins are assumed to have the same function as their homologs, a view that assumes that function is conserved when protein sequences diverge (*37, 38*). This assumption has long been known to be poor in many proteins, including many characterized before the genomic era (*39*). When it fails, it can mislead the biomedical researcher.

A historical analysis based on reconstructed proteins can alert the scientist to the possibility of changing function. When function in a protein family changes, a signature remains within the genome record. Many tools read these signatures, including those that exploit ancestral reconstructions as discussed above. In addition, changing function leaves many traces in the molecular record, including episodes of rapid sequence evolution between reconstructed ancestral sequences, changes in residues that appear to be more easily replaceable (*40, 41*), increases or decreases in the amount of parallel and convergent evolution, and changes in the amount of compensatory covariation (*42*).

For example, the leptin gene in the mouse, when deleted, produced an obese rodent (*43*). The homolog for leptin was found in humans, where it was suggested to be involved in human obesity. This transfer of annotation is consistent with annotation strategies used generally in proteomics (*37, 38*). Examining the reconstructed history of the leptin family in mammals suggested, however, an episode of rapid sequence evolution in the lineages following the divergence of rodents and primates (Fig. 5) (*32, 44*). This suggests that in the small mammals that diverged and gave rise to apes and humans, mutated forms of leptin conferred more fitness than unmutated forms. The functional behavior of leptin must have changed in some way during this time. For the pharmacologist seeking a human antiobesity drug, this suggests (at the very least) that preclinical testing should be done in primate models rather than rodent models (*45*).

Selection of pharmacological models has also been supported by experimental paleobiochemical reconstructions. For example, Chandrasekharan *et al.* resurrected an ancestor of the human and rat chymases (*46*), enzymes that cleave proteins following aromatic residues (phenylalanine, tyrosine, and tryptophan). Human chymase displays high substrate specificity, hydrolyzing a specific peptide bond (Phe^{8}-His^{9}) in angiotensin I to generate angiotensin II, a step in the pathway regulating human blood pressure. Rat chymase-1 does not show this high specificity, instead degrading both angiotensin I and II. The resurrected ancestor of human and rat chymase had high specificity, like the human enzyme, not the lower specificity characteristic of the rodent enzyme (*46*). This result contradicts the hypothesis that serine proteases always evolved from lower to higher specificity. The results also suggest that when one develops drugs targeted against human chymases, primates (not rodents) should be used as experimental models in preclinical assays.

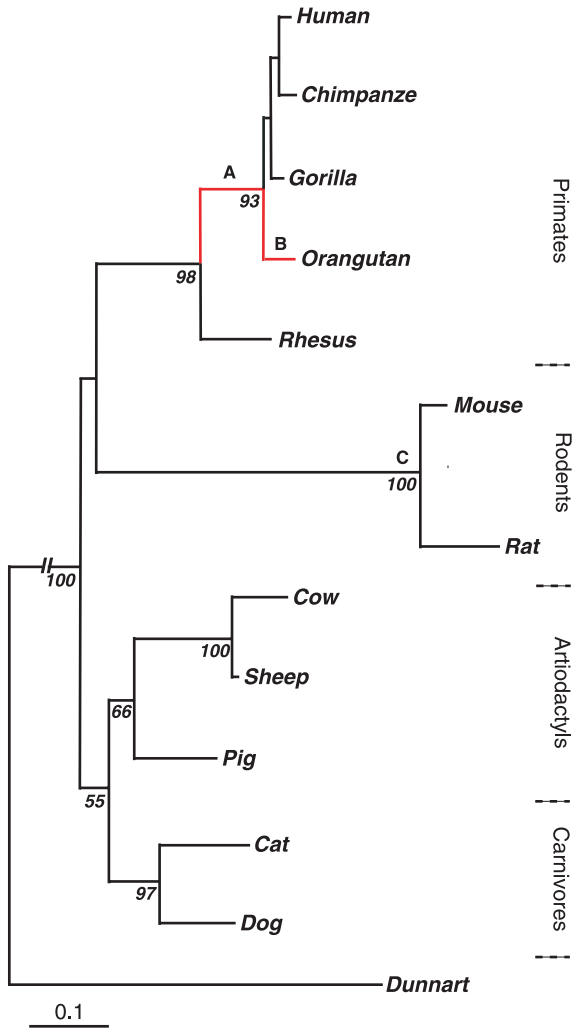

Fig. 5. A model for the evolution of the leptin gene family. The phylogeny is based on numerous morphological and molecular studies and thus represents the consensus for the relationships of the species under study. The numbers below the branches represent the bootstrap values as calculated using PAUP* (version 4.0b5) (*56*), under the maximum likelihood criteria using 100 replicates with replacement for the fast-heuristic search. The following parameters were used: Ti/Tv = 2.3, gamma distribution with alpha = 0.65, and empirical base frequencies (*57*). Branch lengths and the nonsynonymous/synonymous ratios (K_A/K_S) were calculated using PAML (version 3.1) (*58*), incorporating the same parameters as above. Reconstructed evolutionary analyses suggest an episode of rapid sequence evolution of the leptin gene in primitive primates. This branch (A) is highlighted red and has a K_A/K_S ratio equal to 2.18 (9.2 and 1.3 nonsynonymous and synonymous mutations along the branch, respectively). The branch leading to orangutan (B) also contains rapid sequence evolution with a K_A/K_S ratio of 1.23 (4.1 and 1.03 nonsynonymous and synonymous mutations). The high K_A/K_S ratios along these two branches are consistent with positive selection, possibly implying a different function for the primate leptins than for the cenancestral leptin. The branch leading to mouse/rodent (C), however, is highlighted by a K_A/K_S of 0.21. The dunnart (a marsupial) sequence was used to root the tree. The scale bar represents 0.1 substitutions/site/unit evolutionary time.

Planetary Annotation

These examples represent only pieces of a much larger puzzle, one that will have global implications when assembled. Correlation of events in the molecular, paleontological, and geological records, and the molecular dissection of historical events occurring in the past, offers a paradigm to dissect the function of the planetary proteome (*47*). Earth, after all, has only one history. It carries a finite number of species (perhaps between 1 and 2 million). When all of their genomes are sequenced, the planetary proteome will (according to current guesses) be composed of fewer than 10^5 easily recognized modules, independent units of protein sequence evolution (*48–50*). Consequently, one can imagine a comprehensive model of life on Earth, combining paleontology, geology, chemistry, molecular biology, structural biology, systems biology, and genomics, that captures history and function from the molecule to the planet.

Comprehensively annotating the planetary proteome will, of course, be a civilization-wide enterprise, one that has already begun. Various databases assemble the known proteome according to this natural history, including Pfam (*51*), ProDom (*52*), and the MasterCatalog (*53*). The SCOP structure database (*54*), although it does not collect families of sequences, presumably reflects the conservation of fold during sequence divergence. These speed the process of planetary annotation by removing the requirement for a tedious BLAST search and family construction before a historical analysis can begin. Using the MasterCatalog, for example, Liberles *et al.* recently surveyed the entire chordate and embryophyta proteomes, identify-

ing many families that might have undergone adaptive evolution in the past 300 million years (*55*).

A natural annotation of the planetary proteome will require, of course, additions to the paleontological, geological, and molecular records. Already, funding agencies (such as NASA through its Planetary Biology, Exobiology, and Astrobiology programs) are working to improve these records. The consequences will certainly take time to percolate through our educational system, as geologists learn more biology and biologists learn more geology. The past is the key to the present. When we understand where we came from, and how we got here, we understand better who we are. This cannot help but have profound and beneficial impact on health, the environment, and the human condition.

References and Notes

1. L. Pauling, E. Zuckerkandl, *Acta Chem. Scand.* **17** (Suppl. 1), S9 (1963).
2. F. G. R. Liu *et al.*, *Science* **291**, 1786 (2001).
3. W. J. Murphy *et al.*, *Science* **294**, 2348 (2001).
4. R. F. Doolittle, D. F. Feng, S. Tsang, G. Cho, E. Little, *Science* **271**, 470 (1996).
5. S. B. Hedges, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1 (2001).
6. S. A. Bowring *et al.*, *Science* **261**, 1293 (1993).
7. A. C. Wilson, S. S. Carson, T. J. White, *Annu. Rev. Biochem.* **46**, 573 (1997).
8. M. Wallis, *J. Mol. Evol.* **53**, 10 (2001).
9. F. J. Ayala, *Bioessays* **21**, 71 (1999).
10. S. Kumar, S. B. Hedges, *Nature* **392**, 917 (1998).
11. W. H. Li, C. I. Wu, C. C. Luo, *Mol. Biol. Evol.* **2**, 150 (1985).
12. M. D. Caraco, thesis, ETH Zürich, Zürich, Switzerland (2001).
13. M. Fromont-Racine, J. C. Rain, P. Legrain, *Nature Genet.* **16**, 277 (1997).
14. T. Ito *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1143 (2000).
15. M. L. Berbee, J. W. Taylor, *Can. J. Bot.* **71**, 1114 (1993).
16. E. M. Marcotte, M. Pellegrini, M. J. Thompson, T. O. Yeates, D. Eisenberg, *Nature* **402**, 83 (1999).
17. E. M. Marcotte, M. Pellegrini, M. J. Thompson, T. O. Yeates, D. Eisenberg, *Science* **285**, 751 (1999).
18. F. Pazos, A. Valencia, *Protein Eng.* **14**, 609 (2001).
19. D. Dilcher, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7030 (2000).
20. R. T. Bakker, *Nature* **274**, 661 (1978).
21. P. M. Barrett, K. J. Willis, *Biol. Rev.* **76**, 411 (2001).
22. M. Ashburner, *Bioessays* **20**, 949 (1998).
23. H. S. Pereira, D. E. Macdonald, A. J. Hilliker, M. B. Sokolowski, *Genetics* **141**, 263 (1995).
24. D. R. Prothero, *The Eocene-Oligocene Transition: Paradise Lost* (Columbia Univ. Press, NY, 1994).
25. J. A. Wolfe, in *Eocene-Oligocene Climatic and Biotic Evolution*, D. R. Prothero, W. A. Berggen, Eds. (Princeton Univ. Press, Princeton, NJ, 1992), pp. 421–436.
26. J. Zachos, M. Pagani, L. Sloan, E. Thomas, K. Billups, *Science* **292**, 686 (2001).
27. E. A. Barnard, *Nature* **221**, 340 (1969).
28. J. L. Thorne, H. Kishino, J. Felsenstein, *J. Mol. Evol.* **34**, 3 (1992).
29. W. P. Lehman, *Historical Linguistics* (Holt, Rinehart and Winston, New York, 1973).
30. W. Messier, C. B. Stewart, *Nature* **385**, 151 (1997).
31. T. M. Jermann, J. G. Opitz, J. Stackhouse, S. A. Benner, *Nature* **374**, 57 (1995).
32. S. A. Benner, N. Trabesinger-Ruef, D. R. Schreiber, *Adv. Enzyme Regul.* **38**, 155 (1998).
33. L. G. Graddy, thesis, University of Florida, Gainesville, FL (1999).
34. A. M. Dean, G. B. Golding, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3104 (1997).
35. E. A. Gaucher, M. M. Miyamoto, S. A. Benner, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 548 (2001).
36. S. V. Audétat, thesis, ETH Zürich, Zürich, Switzerland (2001).
37. H. Hegyi, M. Gerstein, *Genome Res.* **11**, 1632 (2001).
38. J. S. Fetrow, J. Skolnick, *J. Mol. Biol.* **281**, 949 (1998).
39. S. A. Benner, A. D. Ellington, *CRC Crit. Rev. Biochem.* **23**, 369 (1988).
40. S. A. Benner, *Adv. Enzyme Regul.* **28**, 219 (1989).
41. E. A. Gaucher, X. Gu., M. M. Miyamoto, S. A. Benner, *Trends Biochem. Sci.*, in press.
42. S. A. Benner, D. L. Gerloff, *Adv. Enzyme Regul.* **31**, 121 (1991).
43. Y. Y. Zhang *et al.*, *Nature* **372**, 425 (1994).
44. S. A. Benner, E. A. Gaucher, *Trends Genet.* **17**, 414 (2001).
45. M. Chicurel, *Nature* **404**, 538 (2000).
46. U. M. Chandrasekharan, S. Sanker, M. J. Glynias, S. S. Karnik, A. Husain, *Science* **271**, 502 (1996).
47. L. Margulis, G. R. Fleischaker, *Origins Life Evol. B* **16**, 205 (1986).
48. M. Riley, B. Labedan, *J. Mol. Biol.* **268**, 857 (1997).
49. R. L. Dorit, L. Schoenbach, W. Gilbert, *Science* **250**, 1377 (1990).
50. G. H. Gonnet, M. A. Cohen, S. A. Benner, *Science* **256**, 1443 (1992).
51. A. Bateman *et al.*, *Nucleic Acids Res.* **28**, 263 (2000).
52. F. Corpet, F. Servant, J. Gouzy, D. Kahn, *Nucleic Acids Res.* **28**, 267 (2000).
53. S. A. Benner, S. G. Chamberlin, D. A. Liberles, S. Govindarajan, L. Knecht, *Res. Microbiol.* **151**, 97 (2000).
54. L. Lo Conte *et al.*, *Nucleic Acids Res.* **28**, 257 (2000).
55. D. A. Liberles, D. R. Schreiber, S. Govindarajan, S. G. Chamberlin, S. A. Benner, *Genome Biol.* **2** (no. 8), research0028.1-0028.6 (2001).
56. D. L. Swofford, *Phylogenetic Analysis Using Parsimony (*and other methods)* (Sinauer Associates, Sunderland, MA, 1998).
57. E. A. Gaucher, thesis, University of Florida, Gainesville, FL (2001).
58. Z. Yang, *Comput. Appl. Biosci.* **15**, 555 (1997).

59. We are indebted to funding from the NASA Exobiology Program, the NASA Astrobiology Institute, the NIH (GM 54075), the National Research Council (an associateship to E.A.G.), and the Agouron Foundation. We are indebted to S. Chamberlin of EraGen Biosciences for many helpful discussions and for access to the MasterCatalog, EraGen's interpretive proteomics platform.

REVIEW: NEUROSCIENCE

Axonal Self-Destruction and Neurodegeneration

Martin C. Raff,* Alan V. Whitmore, John T. Finn†

Neurons seem to have at least two self-destruct programs. Like other cell types, they have an intracellular death program for undergoing apoptosis when they are injured, infected, or not needed. In addition, they apparently have a second, molecularly distinct self-destruct program in their axon. This program is activated when the axon is severed and leads to the rapid degeneration of the isolated part of the cut axon. Do neurons also use this second program to prune their axonal tree during development and to conserve resources in response to chronic insults?

Much effort is being devoted to understanding the nature of neuronal cell death in various neurodegenerative diseases such as motor neuron disease, glaucoma, and Alzheimer, Parkinson, and Huntington diseases (*1*–*5*). It may be, however, that neuronal death in these diseases occurs too late to be clinically important. Degeneration of the neuron's long process—the axon—often precedes the death of the cell body and may make a more important contribution to the patient's disability.

Here, we discuss some examples of axonal degeneration in disease and in normal development. We consider one neurodegenerative disease in which axonal degeneration, rather than neuronal death, seems to be responsible for clinical progression and death. We review the evidence that axonal degeneration may occur through a local self-destruct program, which is distinct from the proteolytic program that mediates apoptosis (programmed cell death). We speculate that the same axonal self-destruct program may be used by developing neurons to eliminate unwanted axonal branches, and by unhealthy neurons to eliminate an injured axon or to disconnect from their postsynaptic targets to conserve resources.

MRC Laboratory for Molecular Cell Biology and Cell Biology Unit and the Biology Department, University College London, London WC1E 6BT, UK.

*To whom correspondence should be addressed. E-mail: m.raff@ucl.ac.uk

†Present address: Systems Research Department, Sandia National Laboratories, Post Office Box 969, Mail Stop 9201, Livermore, CA 94551, USA.