



# A calibrated chronology of biochemistry reveals a stem line of descent responsible for planetary biodiversity

Gustavo Caetano-Anollés<sup>1\*</sup>, Jay E. Mittenenthal<sup>2</sup>, Derek Caetano-Anollés<sup>1</sup> and Kyung Mo Kim<sup>3</sup>

<sup>1</sup> Evolutionary Bioinformatics Laboratory, Department of Crop Sciences, University of Illinois, Urbana, IL, USA

<sup>2</sup> Department of Cell and Developmental Biology, University of Illinois, Urbana, IL, USA

<sup>3</sup> Microbial Resource Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, South Korea

## Edited by:

Michel Laurin, CNRS, UMR 7207,  
Muséum National d'Histoire  
Naturelle, France

## Reviewed by:

Alex Piron, The George Washington  
University, USA  
Michel Laurin, CNRS, UMR 7207,  
Muséum National d'Histoire  
Naturelle, France

## \*Correspondence:

Gustavo Caetano-Anollés,  
Evolutionary Bioinformatics  
Laboratory, Department of Crop  
Sciences, University of Illinois at  
Urbana-Champaign, 332 NSRC, 1101  
W. Peabody Drive, Urbana, IL  
61801, USA  
e-mail: gca@illinois.edu

Time-calibrated phylogenomic trees of protein domain structure produce powerful chronologies describing the evolution of biochemistry and life. These timetrees are built from a genomic census of millions of encoded proteins using models of nested accumulation of molecules in evolving proteomes. Here we show that a primordial stem line of descent, a propagating series of pluripotent cellular entities, populates the deeper branches of the timetrees. The stem line produced for the first time cellular grades ~2.9 billion years (Gy)-ago, which slowly turned into lineages of superkingdom Archaea. Prompted by the rise of planetary oxygen and aerobic metabolism, the stem line also produced bacterial and eukaryal lineages. Superkingdom-specific domain repertoires emerged ~2.1 Gy-ago delimiting fully diversified Bacteria. Repertoires specific to Eukarya and Archaea appeared 300 millions years later. Results reconcile reductive evolutionary processes leading to the early emergence of Archaea to superkingdom-specific innovations compatible with a tree of life rooted in Bacteria.

**Keywords:** structure, phylogenetic analysis, molecular clock, protein folds, protein evolution

A stem line of descent is a primordial propagating series of pluripotent cellular entities. It is believed that such a remarkable backbone of the cellular world resulted in sequential cellular lineage spin-offs (Wang et al., 2007), very much as modern differentiated cell types arise from less differentiated embryonic or adult stem cells. The axioms of evolution supporting the genealogy of life seeded by the stem line demand that tree-like ensembles of lineages nested within each other populate the biological world (Wiley, 1975). The molecular, cellular, and organismal entities (taxa) that unfold in the nested lineages are the subject of gradual change, fulfilling spatiotemporal continuity and benefitting from non-vertical transfer of information. Taxa also retain features (characters) that are refractory to change and are preserved through genealogical descent. These characters allow dissection of the history of life and the generation of evolutionary chronologies embodied in time-calibrated phylogenetic trees (timetrees). As we will show, characters of this type, such as the highly conserved atomic structure of macromolecules (Caetano-Anollés et al., 2009), can also inform about the ancient history of the stem line and its makeup.

When taxa are organisms and characters are organismal component parts, trees represent explicit statements of organismal history. Once crucial divergences are time-calibrated with molecular, fossil and other evidence, the nested hierarchies represent “timetrees of life” (ToLs) (Donoghue and Benton, 2007; Laurin, 2012). The leaves of these traditional trees depict present-day organisms (sometimes extinct taxa representing species, genera, or families) and their evolution is viewed through the lens of

the conserved features that are studied; for example, a tree of  $\alpha$ -proteobacteria can be viewed through evolution of their ribosomal RNA molecules. Conversely, when taxa are the component parts themselves, the nested hierarchies describe a different kind of timetree that explores the chronology of component innovation of the organismal system. In particular, “timetrees of domains” (ToDs) are powerful trees that harbor leaves representing protein domain structures, their evolution viewed through the lens of their genomic abundance (Wang et al., 2011). Here we show that these time-calibrated trees uncover remarkable patterns of origin and diversification of biochemistry, striking evolutionary patterns supporting the ancestral stem line, and the time of emergence of planetary biodiversity.

## PHYLOGENOMIC RETRODICTION UNCOVERS BIOCHEMICAL HISTORY

Biological “modules” that are sufficiently conserved are particularly useful characters for phylogenetic analysis. Modules are sets of component parts that interact more tightly with each other than with other parts of the system. They can be molecular (e.g., protein domains; Murzin et al., 1995) or of other kinds (e.g., developmental; Laurin, 2014). These cohesive units have the ability to diversify in the many different contexts of the cell, adding to biodiversity (Mittenenthal et al., 2012). The number of module types increases in evolution; their growing numbers also distribute in nested manner in the nested lineages of the trees. Useful molecular modules include amino acid monomers in protein chains, sets of stabilizing H-bonding interactions in RNA

molecules, structural motifs in folded proteins, or protein domain structures in proteomes. Since modules can be viewed as determinants of levels of biological organization, timetrees of modules that are part of the molecular and cellular makeup of the cells can provide useful information about biochemical history.

Evolved macromolecules are generally endowed with a finite conformational ensemble of 3-dimensional atomic structures (e.g., Fontana, 2002; Schuster, 2010). These folding conformations are highly dynamic and materialize for significant periods of time, enough to hold molecular functions that are advantageous to the cells. Consequently, their ensembles represent “living fossils” that have retained considerable history of their molecular past (Caetano-Anollés et al., 2009). They can be mined with modern bioinformatics tools of phylogenetic reconstruction.

We have shown that the natural history of protein domain structures and their associated functions can be directly inferred from ToDs, comb-like trees that harbor domains as taxa (Caetano-Anollés and Caetano-Anollés, 2003). ToDs are built from a genomic census of protein domain structure (suitably defined) in the proteomes of thousands of genomes using the standard tools of cladistics analysis. Domains are the main structural, functional and evolutionary modules of proteins that are highly conserved and are recurrently arranged in multidomain proteins (Wang and Caetano-Anollés, 2009). The genomic census of folded structures can be defined using the different levels of structural abstraction of the accepted protein domain classification gold standards, the SCOP (Murzin et al., 1995) or CATH (Orengo et al., 1997) databases. **Figure 1** shows a data matrix (array) of domain abundance in the proteomes of 420 free-living organisms belonging to the three superkingdoms of cellular life, Archaea, Bacteria, and Eukarya. The analysis excludes the history of viruses, which can be found elsewhere (Nasir et al., 2012). The phylogenetic matrix was visualized as a heat map, a graphical representation of the data in which values of domain abundance are represented with colors. Domains in the genomic census were defined at the fold superfamily (FSF) level of the SCOP hierarchical classification. FSFs unify a diversity of protein sequences that share a 3-dimensional fold structure, related molecular functions and a common evolutionary origin. In the data matrix, the proteomes of organisms (rows of the matrix) were ordered according to a rooted and ladderized ToL. Similarly, domains in proteomes (columns) were ordered according to the relative ages of FSFs (*nd*), which were derived directly from a ToD. The rooted ToL and ToD trees are shown to the left and below the heat map and were built using maximum parsimony from the data matrix and its transposed derivative, respectively.

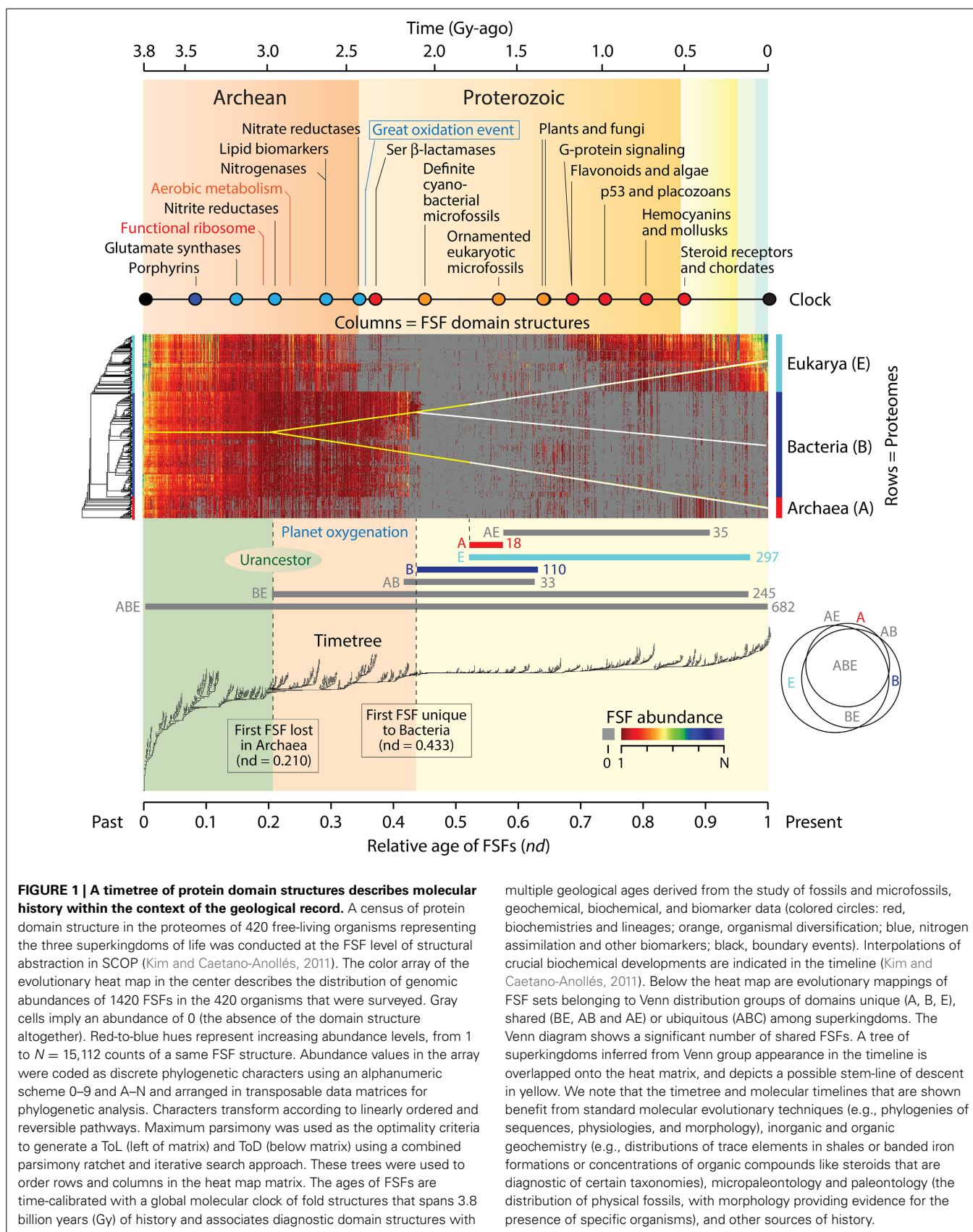
We note that three properties enable timetree retrodictions from domain abundance data (**Figure 1**): (i) *The trees are rooted*: Rooted ToDs are built by using a process model that considers that the most abundant and widely distributed domain structures are of ancient origin (Caetano-Anollés and Caetano-Anollés, 2003). The model confers polarity (distinction between ancestral and derived states) to a transformation series of ordered multi-state phylogenetic characters that describe increases and decreases of FSF abundances in proteomes. This polarization roots the trees without invoking outgroup taxa or other external

assumptions and can be validated by a number of criteria (Kim et al., 2013). It also complies with Weston’s generality criterion (Weston, 1994), which is supported by homology in nested patterns and additive phylogenetic change and roots the trees with the minimum number of assumptions. Using the Lundberg method, the root is identified by attaching a hypothetical ancestor that is defined by the polarization model to an optimal unrooted tree in a most-parsimonious manner (Lundberg, 1972). (ii) *Chronologies are inferred directly from the trees*: Chronologies cannot be inferred directly from rooted trees that tend to follow stochastic or null branching processes of change, i.e., that are relatively well balanced. In these cases, time calibrations for origins of clades are achieved for example by the use of fossil data in ToLs. In contrast, when rooted trees follow semi-punctuated evolutionary processes responsible for accelerated change during divergence (Venditti and Pagel, 2010), they are highly unbalanced and pectinate in appearance. This is the case of ToDs, in which splitting of lineages depends on an evolving “heritable” trait (Heard, 1996), the gradual accumulation of structural variants of domains in lineages and the semipunctuated discovery of new domain structures. Chronologies can be inferred directly from these imbalanced trees by calculating a “node distance” (*nd*), the relative number of internal nodes from the root to a leaf of the tree. (iii) *Phylogenetic statements are linked to the geological record through time calibration points*: A global molecular clock of domain structures establishes a significant linear relationship between the age of domains and the geological record (Wang et al., 2011). Thus, FSF domain structures diagnostic of biomarkers and geomarkers provide the time calibration points.

The time of first appearance of a domain structure at FSF level in the chronology records the time of the origin of that FSF. Consequently, the chronology of FSFs should be viewed as a timeline of molecular innovation portraying the gradual rise of modern biochemistry (Caetano-Anollés et al., 2009). For example, we have used chronologies of these kinds to trace the origin and evolution of metabolic networks (Caetano-Anollés et al., 2007; Kim et al., 2012; Caetano-Anollés and Caetano-Anollés, 2013), study the rise of translation and the genetic code (Caetano-Anollés et al., 2011, 2012, 2013), uncover the coevolutionary history of the ribosome (Harish and Caetano-Anollés, 2012), explore the evolution of metallomes and biological metal utilization (Dupont et al., 2010), unfold the natural history of biocatalytic mechanisms (Nath et al., 2014), study the evolutionary dynamics of gain and loss of domains (Nasir et al., 2014), and determine the makeup of the common ancestor of life (Kim and Caetano-Anollés, 2011).

## PATTERNS OF DOMAIN ABUNDANCE REVEAL A MEGAORGANISMIC STEM LINE OF DESCENT

The heat maps of **Figure 1** overlap domain abundance (and occurrence) on the data matrix used for timetree generation. They describe the reuse and spread of molecular innovations in the proteomes of the modern protein world. Older FSFs are expected to be more abundant and widely distributed (Wang et al., 2007, 2011). Indeed, the most abundant and widely spread FSFs that are universally distributed appear at the base of the



chronology, within the first 0.1 Gy of protein history. These structures are part of an initial and universal core responsible for the primordial metabolic, structural, and cellular functions (Danchin et al., 2007; Wang et al., 2007), which we propose represents the most ancient repertoire of the stem line that gave rise to the molecular and organismal biodiversity of the planet. This stem line was probably embodied in a megaorganism, in the sense of a modern fusion-driven syncytium and/or a fission-driven coenocyte (such as the plasmodial slime molds; Egel, 2012), which preceded modern organisms and lineages. The entity probably resembled the multiphenotypical precells of Kandler (1994) that later seeded the concept of a communal ancestor (Woese, 2002). The FSF repertoire was most likely quasispecies-like and fluid, i.e., a cloud of genotypes and phenotypes changing at high rate (Seufferheld and Caetano-Anollés, 2013). It was cellular and harbored a molecular makeup that was in part associated with archaic membranes. Iterative phylogenetic character state reconstruction of the proteome of the urancestral line suggests it contained 303–507 domains apportioned into 70–152 FSFs at the time of the appearance of first lineages (Kim and Caetano-Anollés, 2011). Remarkably, conical laminated morphologies of probable biological origin (stromatolites) have been identified in the 3.4 Gy-old Strelley Pool formation of Western Australia (Allwood et al., 2006) and may contain microfossils of sulfur-metabolizing cells (Wacey et al., 2011). Our analysis now provides molecular backing to the biogenic interpretation of these rock microstructures.

While the proteomic repertoire of the stem line was 3–5 times smaller than that of standard free-living organisms, its makeup continued to expand. As time progressed in the timeline and new FSFs were uncovered, the differential accumulation of FSFs in proteomes makes evident the evolutionary expansion/reduction of proteome repertoires. It shows that a number of FSFs were gradually lost (or never gained) in archaeal organisms. These early diversification patterns suggest the emergence of archaeal grades by reductive evolution and the very early rise of Archaea as the first superkingdom of diversified life (Wang et al., 2007; Caetano-Anollés et al., 2014). Grades (*sensu* Huxley, 1958) are variants of the stem line in active transition but unified by the same level of physiological complexity stemming from the FSF repertoire (Caetano-Anollés et al., 2014). We note that complete loss of an FSF in a proteome is most likely to occur soon after the origin of the FSF, when there are few copies of it encoded in the genome. Since genomic abundances of each FSF increase in evolution as genes duplicate and diverge or are created *de novo*, it is increasingly less likely for domain loss to occur later on in evolution. This is especially so if loss occurs in all proteomes of a superkingdom, such as the first FSF that was lost completely in all lineages of Archaea ~2.9 Gy ago. It is also noteworthy that when domain abundances in FSFs increase, the probability of recruitment of FSF variants to perform different functions in different cellular contexts also increases. This fosters domain innovation, domain interactions, multidomain proteins, and the formation of cellular complexes. Thus, the age of an FSF is just the lower bound on its use as a module—maybe a bound unlikely to be attained—since modularity unfolds gradually in evolution (e.g., domains as recombining units; Wang and Caetano-Anollés, 2009).

## THE RISE OF PLANETARY BIODIVERSITY BEGAN ~2.9 Gy AGO AND MATERIALIZED ~2.1 Gy AGO FOLLOWING THE “CRYSTALLIZATION” OF MODERN CELLULAR MODULES

The same reductive evolutionary trend of differential accumulation of FSFs different from the stem line that starts in Archaea ~2.9 Gy ago is seen for the first time ~2.8 Gy ago in Bacteria and Eukarya (their common ancestral line), revealing further weakening of the stem line in favor of grades (Figure 1). This becomes especially evident 2.45 Gy ago, the time of the Great Oxygenation Event (GOE) of the planet (Sessions et al., 2009). Abundances of new FSFs in Eukarya drop precipitously at this time in the heat map, while those in Bacteria were maintained for an additional 3–4 million years (Figure 1). The fact that these sharp reductive changes manifest with prior and time-localized increases of FSF abundances in Eukarya 2.6–2.9 Gy ago and in Bacteria ~2.4 Gy ago is remarkable. It suggests a complex dynamics of domain growth in FSFs unfolding close to the transition point (threshold or crystallization; *sensu* Woese, 2002) between the megaorganismic precells and grades and the rise of modern cells and lineages. This transition is marked by the appearance of the first superkingdom-specific FSFs in Bacteria ~2.1 Gy ago and coincides with the age of the oldest unambiguous microfossils of cyanobacteria (matching the genus *Archaeoellipsoides*) that are preserved in 1.5–2.1 Gy old cherts throughout the world (Tomitani et al., 2006). We note that the appearance of these FSFs should be regarded as unequivocal. Their origin cannot result from transfer or massive loss of structures in other superkingdoms.

## LIKELY CULPRITS OF PLANETARY BIODIVERSITY: OXYGEN AND AEROBIC METABOLISM

Timetrees suggest that aerobic metabolism and planetary oxygenation appeared concurrently ~2.9 Gy ago (Wang et al., 2011), ~400 million years before the GOE, and that the Mn catalase enzyme was the ultimate culprit of oxygen production (Kim et al., 2012). Chemoinformatic dissection of the chemical space of metabolites showed that planetary oxygen shaped the chemical makeup of metabolites in aerobic metabolic networks, expanding the structural and chemical space with 130 new molecular scaffolds of the 335 total set (Jiang et al., 2012). Oxygen also made metabolites more rigid and hydrophobic, when these were analyzed with a range of chemical property descriptors (e.g., rotatable bond counts, polar molecular volume, hydrophobic fragment count). These properties prompted cellular diversity. For example, the polar makeup of steroids influences endo/exocytosis and transmembrane trafficking, poisoning the transition to multicellularity (Summons et al., 2006). It is also likely that planetary oxygenation resulted in extinction of a significant part of anaerobic biodiversity, enabling novel organismal radiations and unfolding adaptations in response to the new noxious environments, which are currently reflected in links between metabolism and ontogenesis (Herkovits, 2006). We therefore propose that both oxygen and aerobic metabolism enhanced the complexity and diversity of cellular organization.

## CONCLUSION

Phylogenomics makes explicit the differential evolutionary accumulation of domain structures in evolving proteomes.

Evolutionary patterns suggest that the primordial stem line produced cellular grades 2.9 Gy ago, which slowly turned into archaeal and eukaryal lineages and into bacterial lineages at more accelerated pace. This differential behavior reconciles the early origin of Archaea inferred from molecular structures and functions with the canonical rooting of the ToL sometimes revealed by standard sequence analysis. The important realization that supports these findings is the continuous growth of the protein world and the fact that the complete loss of a domain structure in a proteome occurs more likely soon after its origin in the timeline, at a time when there are few domain copies encoded in the genome and recruitment is incipient.

## REFERENCES

- Allwood, A. C., Walter, M. R., Kamber, B. S., Marshall, C. P., and Burch, I. W. (2006). Stromatolite reef from the early Archaean era of Australia. *Nature* 441, 714–718. doi: 10.1038/nature04764
- Caetano-Anollés, D., Kim, K. M., Mittenthal, J. E., and Caetano-Anollés, G. (2011). Proteome evolution and the metabolic origins of translation and cellular life. *J. Mol. Evol.* 72, 14–33. doi: 10.1007/s00239-010-9400-9
- Caetano-Anollés, G., and Caetano-Anollés, D. (2003). An evolutionarily structured universe of protein architecture. *Genome Res.* 13, 1563–1571. doi: 10.1101/gr.1161903
- Caetano-Anollés, G., Kim, H. S., and Mittenthal, J. E. (2007). The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9358–9363. doi: 10.1073/pnas.0701214104
- Caetano-Anollés, G., Kim, K. M., and Caetano-Anollés, D. (2012). The phylogenomic roots of modern biochemistry: origins of proteins, cofactors and protein biosynthesis. *J. Mol. Evol.* 74, 1–34. doi: 10.1007/s00239-011-9480-1
- Caetano-Anollés, G., Nasir, A., Zhou, K., Caetano-Anollés, D., Mittenthal, J. E., Sun, F. -J., et al. (2014). Archaea: the first domain of diversified life. *Archaea* 2014:590214. doi: 10.1155/2014/590214
- Caetano-Anollés, G., Wang, M., and Caetano-Anollés, D. (2013). Structural phylogenomics retrodicts the origin of the genetic code and uncovers the evolutionary impact of protein flexibility. *PLoS ONE* 8:e72225. doi: 10.1371/journal.pone.0072225
- Caetano-Anollés, G., Wang, M., Caetano-Anollés, D., and Mittenthal, J. E. (2009). The origin, evolution and structure of the protein world. *Biochem. J.* 417, 621–637. doi: 10.1042/BJ20082063
- Caetano-Anollés, K., and Caetano-Anollés, G. (2013). Structural phylogenomics reveals gradual evolutionary replacement of abiotic chemistries by protein enzymes in purine metabolism. *PLoS ONE* 8:e59300. doi: 10.1371/journal.pone.0059300
- Danchin, A., Fang, G., and Noria, S. (2007). The extant core bacterial proteome is an archive of the origin of life. *Proteomics* 7, 875–889. doi: 10.1002/pmic.200600442
- Donoghue, P. C. J., and Benton, M. J. (2007). Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends Ecol. Evol.* 22, 424–431. doi: 10.1016/j.tree.2007.05.005
- Dupont, C. L., Butcher, A., Valas, R. E., Bourne, P. E., and Caetano-Anollés, G. (2010). History of biological metal utilization inferred through phylogenomic analysis of protein structures. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10567–10572. doi: 10.1073/pnas.0912491107
- Egel, R. (2012). Primal eukaryogenesis: on the communal nature of precellular states, ancestral to modern life. *Life* 2, 170–212. doi: 10.3390/life2010170
- Fontana, W. (2002). Modelling ‘evo-devo’ with RNA. *Bioessays* 24, 1164–1177. doi: 10.1002/bies.10190
- Harish, A., and Caetano-Anollés, G. (2012). Ribosomal history reveals origins of modern protein synthesis. *PLoS ONE* 7:e32776. doi: 10.1371/journal.pone.0032776
- Heard, S. B. (1996). Patterns of phylogenetic tree balance with variable or evolving speciation rates. *Evolution* 50, 2141–2148. doi: 10.2307/2410685
- Herkovits, J. (2006). Evoecotoxicology: environmental changes and life features development during the evolutionary process – the record of the past at developmental stages of living organisms. *Environ. Health Perspect.* 114, 1139–1142. doi: 10.1289/ehp.8633
- Huxley, J. S. (1958). Evolutionary processes and taxonomy with special reference to grades. *Uppsala Universitets Årsskrift* 6, 21–39.
- Jiang, Y.-Y., Kong, D.-K., Qin, T., Li, X., Caetano-Anollés, G., and Zhang, H.-Y. (2012). The impact of oxygen on metabolic evolution: a cheminformatic investigation. *PLoS Comput. Biol.* 8:e1002426. doi: 10.1371/journal.pcbi.1002426
- Kandler, O. (1994). Cell wall biochemistry and three-domain concept of life. *Syst. Appl. Microbiol.* 16, 501–509. doi: 10.1016/S0723-2020(11)80319-X
- Kim, K. M., and Caetano-Anollés, G. (2011). The proteomic complexity and rise of the primordial ancestor of diversified life. *BMC Evol. Biol.* 11:140. doi: 10.1186/1471-2148-11-140
- Kim, K. M., Nasir, A., and Caetano-Anollés, G. (2013). The importance of using realistic evolutionary model for retrodicting proteomes. *Biochimie* 99, 129–137. doi: 10.1016/j.biochi.2013.11.019
- Kim, K. M., Qin, T., Jiang, Y.-Y., Xiong, M., Caetano-Anollés, D., Zhang, H.-Y., et al. (2012). Protein domain structure uncovers the origin of aerobic metabolism and the rise of planetary oxygen. *Structure* 20, 67–76. doi: 10.1016/j.str.2011.11.003
- Laurin, M. (2012). Recent progress in paleontological methods for dating the Tree of Life. *Front. Genet.* 3:130. doi: 10.3389/fgene.2012.00130
- Laurin, M. (2014). Assessment of modularity in the urodele skull: an exploratory analysis using ossification sequence data. *J. Exp. Zool. B Mol. Dev. Evol.* doi: 10.1002/jez.b.22575. [Epub ahead of print].
- Lundberg, J. G. (1972). Wagner networks and ancestors. *Syst. Biol.* 21, 398–413. doi: 10.1093/sysbio/21.4.398
- Mittenthal, J. E., Caetano-Anollés, D., and Caetano-Anollés, G. (2012). Biphasic patterns of diversification and the emergence of modules. *Front. Genet.* 3:147. doi: 10.3389/fgene.2012.00147
- Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540. doi: 10.1016/S0022-2836(05)80134-2
- Nasir, A., Kim, K. M., and Caetano-Anollés, G. (2012). Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. *BMC Evol. Biol.* 12:156. doi: 10.1186/1471-2148-12-156
- Nasir, A., Kim, K. M., and Caetano-Anollés, G. (2014). Global patterns of protein domain gain and loss in superkingdoms. *PLoS Comput. Biol.* 10:e1003452. doi: 10.1371/journal.pcbi.1003452
- Nath, N., Mitchell, J. B. O., and Caetano-Anollés, G. (2014). The natural history of biocatalytic mechanisms. *PLoS Comput. Biol.* 10:e1003642. doi: 10.1371/journal.pcbi.1003642
- Orengo, C. A., Michie, A., Jones, S., Jones, D. T., Swindells, M., and Thornton, J. M. (1997). CATH—a hierarchical classification of protein domain structures. *Structure* 5, 1093–1109. doi: 10.1016/S0969-2126(97)00260-8
- Schuster, P. (2010). “Genotypes and phenotypes in the evolution of molecules,” in *Evolutionary Bioinformatics and Systems Biology*, ed G. Caetano-Anollés (Hoboken, NJ: John Wiley & Sons), 123–152.
- Sessions, A. L., Doughty, D. M., Welander, P. V., Summons, R. E., and Newman, D. K. (2009). The continuing puzzle of the great oxidation event. *Curr. Biol.* 19, R567–R574. doi: 10.1016/j.cub.2009.05.054
- Seufferheld, M., and Caetano-Anollés, G. (2013). Phylogenomics supports a cellularly structured urancestor. *J. Mol. Microbiol. Biotechnol.* 23, 178–191. doi: 10.1159/000346552
- Summons, R. E., Bradley, A. S., Jahnke, L. L., and Waldbauer, J. R. (2006). Steroids, triterpenoids and molecular oxygen. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 951–968. doi: 10.1098/rstb.2006.1837
- Tomitani, A., Knoll, A. H., Cavanaugh, C. M., and Ohno, T. (2006). The evolutionary diversification of cyanobacteria: molecular, phylogenetic and paleontological perspectives. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5442–5447. doi: 10.1073/pnas.0600999103
- Venditti, C., and Pagel, M. (2010). Speciation as an active force in promoting genetic variation. *Trends Ecol. Evol.* 25, 14–20. doi: 10.1016/j.tree.2009.06.010
- Wacey, D., Kilburn, M. R., Saunders, M., Cliff, J., and Brasier, M. D. (2011). Microfossils of sulphur-metabolizing cells in 3.40 billion-year-old rocks of Western Australia. *Nat. Geosci.* 4, 698–702. doi: 10.1038/ngeo1238
- Wang, M., and Caetano-Anollés, G. (2009). The evolutionary mechanics of domain organization in proteomes and the rise of modularity in the protein world. *Structure* 17, 66–78. doi: 10.1016/j.str.2008.11.008

- Wang, M., Jiang, Y.-Y., Kim, K. M., Qu, G., Ji, H.-F., Mittenthal, J. E., et al. (2011). A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation. *Mol. Biol. Evol.* 28, 567–582. doi: 10.1093/molbev/msq232
- Wang, M., Yafremava, L. S., Caetano-Anollés, D., Mittenthal, J. E., and Caetano-Anollés, G. (2007). Reductive evolution of architectural repertoires in proteomes and the birth of the tripartite world. *Genome Res.* 17, 1572–1585. doi: 10.1101/gr.6454307
- Weston, P. H. (1994). “Methods for rooting cladistic trees,” in *Models in Phylogeny Reconstruction*, eds D. J. Siebert, R. W. Scotland, and D. M. Williams (Oxford: Oxford University Press), 125–155.
- Wiley, E. O. (1975). Karl R. Popper, systematics, and classification: a reply to Walter Bock and other evolutionary taxonomists. *Syst. Zool.* 24, 233–243. doi: 10.2307/2412764
- Woese, C. R. (2002). On the evolution of cells. *Proc. Natl. Acad. Sci. U.S.A.* 99, 8742–8747. doi: 10.1073/pnas.132266999
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 June 2014; accepted: 18 August 2014; published online: 11 September 2014.

Citation: Caetano-Anollés G, Mittenthal JE, Caetano-Anollés D and Kim KM (2014) A calibrated chronology of biochemistry reveals a stem line of descent responsible for planetary biodiversity. *Front. Genet.* 5:306. doi: 10.3389/fgene.2014.00306

This article was submitted to *Evolutionary and Population Genetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Caetano-Anollés, Mittenthal, Caetano-Anollés and Kim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.