Free Radical Biology and Medicine 140 (2019) 260–269

Contents lists available at ScienceDirect

Free Radical Biology and Medicine

journal homepage: www.elsevier.com/locate/freeradbiomed

How to resurrect ancestral proteins as proxies for ancient biogeochemistry

Amanda K. Garcia[a], Betül Kaçar[a,b,*]

[a] Department of Molecular and Cell Biology, University of Arizona, Tucson, AZ, 85721, USA
[b] Department of Astronomy and Steward Observatory, University of Arizona, Tucson, AZ, 85721, USA

ARTICLE INFO

Keywords:
Ancestral sequence reconstruction
Great oxidation event
Nitrogenase
Paleophenotype
Phylogenetic uncertainty
RuBisCO
Biosignatures

ABSTRACT

Throughout the history of life, enzymes have served as the primary molecular mediators of biogeochemical cycles by catalyzing the metabolic pathways that interact with geochemical substrates. The byproducts of enzymatic activities have been preserved as chemical and isotopic signatures in the geologic record. However, interpretations of these signatures are limited by the assumption that such enzymes have remained functionally conserved over billions of years of molecular evolution. By reconstructing ancient genetic sequences in conjunction with laboratory enzyme resurrection, preserved biogeochemical signatures can instead be related to experimentally constrained, ancestral enzymatic properties. We may thereby investigate instances within molecular evolutionary trajectories potentially tied to significant biogeochemical transitions evidenced in the geologic record. Here, we survey recent enzyme resurrection studies to provide a reasoned assessment of areas of success and common pitfalls relevant to ancient biogeochemical applications. We conclude by considering the Great Oxidation Event, which provides a constructive example of a significant biogeochemical transition that warrants investigation with ancestral enzyme resurrection. This event also serves to highlight the pitfalls of facile interpretation of paleophenotype models and data, as applied to two examples of enzymes that likely both influenced and were influenced by the rise of atmospheric oxygen – RuBisCO and nitrogenase.

1. Introduction

Exploration of the interface between ancient biological and geochemical systems requires an understanding of how both systems functioned in the past. Whereas uniformitarian assumptions can be credibly applied to much of the geologic record owing to the immutability of the physico-chemical processes that produce and preserve geologic materials, the limits of such assumptions are quickly reached with biological organisms and structures. Modern organisms and their genomes are the products of oftentimes contingent evolutionary trajectories, and it is not always clear to what extent contemporary biological characteristics represent those of their ancient predecessors. Thus, it is helpful to consider the ways by which extant organisms and molecular systems may be modified, such that they may reveal ancestral functions relevant to ancient biogeochemical processes.

Enzymes have served a prime role as molecular mediators of biological and geochemical cycles. Their activities have shaped the chemical and isotopic compositions of preserved materials, interpreted as biosignatures in the geologic record. Though admittedly a reductionist perspective, enzymes can offer a practicable target for molecular resurrection studies. By relating ancestral enzyme properties to preserved biosignatures, we may constrain aspects of the ancient environment in which they functioned, as well as identify instances within biomolecular evolutionary trajectories that may be tied to significant biogeochemical events.

The Great Oxidation Event (ca. 2.1–2.4 Ga [1,2]) is one such example. The progressive oxygenation of the Earth's surface environment undoubtedly imposed new oxidative stresses on life at both organismal and molecular levels [3–6] as well as yielded a new repertoire of available trace nutrients and enzymatically-catalyzed metabolic pathways [7–11]. These effects may be resolvable in properties of reconstructed ancient enzymes. By use of paleogenetics, whereby the sequences of ancestral biomolecules are statistically inferred and resurrected in the laboratory, connections may be made between the global-scale effects of the Great Oxidation Event, organism-scale metabolism, and biomolecule-scale catalysis.

Recent studies have begun using inferred ancestral enzymes to explore the interaction between atmospheric oxygenation and the evolution of biomolecular systems over geologic timescales [12–14]. The continued expansion of such investigations may provide promising insights into the ancient history of biogeochemical cycling. However, uncertainties associated with the phylogenetic methods that underlie

* Corresponding author. Department of Molecular and Cellular Biology, Life Sciences South Building, 1007 E. Lowell Street, PO BOX 210106, Tucson, AZ, 85721, USA.
E-mail address: betul@arizona.edu (B. Kaçar).

https://doi.org/10.1016/j.freeradbiomed.2019.03.033
Received 17 September 2018; Received in revised form 11 February 2019; Accepted 26 March 2019
Available online 02 April 2019

paleogenetic inference can impact the reliability of reconstructed molecular sequences, preventing accurate assessments of ancestral enzyme function.

Here we survey sources of phylogenetic bias in the context of paleogenetics and review recent applications relevant to Earth systems that may provide direction for future biogeochemical studies. We then discuss how we can meaningfully extend these applications beyond single-molecule resurrections toward the assessment of ancestral molecular systems expressed within modern microbes.

2. A uniformitarian approach to Earth's biomolecular history

All forms of extant life have genes that broadly define the contours of relatedness between different organisms. Closely-related organisms typically have genes and gene sequences that are more comparable to one another than distantly-related organisms [15]. Some genes are peripheral to organismal survival and may exhibit rapid rates of mutation without affecting organismal fitness; others are essential to an organism's metabolic or reproductive success and are thus more likely to be conserved [16–19]. In certain cases, essential genes have preserved functionality and possibly even significant sequence identity with the earliest organisms that inhabited the Earth billions of years ago [20,21]. These two end members of sequence variability form a space in which studies into life's genetic past may be conceived, conducted and compared with the geologic record. However, the mapping of mutable biological or biochemical similarities is arguably more arduous than extending immutable physical and chemical relationships across geologic timescales.

A major reason for this difficulty rests on complications that arise when connecting gene sequence information with biomolecular and organismal phenotype. As a typical example, consider a gene that codes for a protein with a critical catalytic role. Within that protein sequence, some portions will contribute more to the protein's function than others. These critical portions are commonly sequence sites associated with a protein's folding, structure, catalytic metal-cofactor binding centers, or protein-protein interactions (i.e., the physical contacts between proteins that are determined by their structure and biochemistry) [22,23]. However, a seemingly inconsequential mutation in a non-critical part of the protein might unexpectedly enable a novel function; similarly, a series of otherwise individually innocuous mutations introduced throughout the protein might lead to a novel function or a more efficient catalytic property. In some cases, very few amino acid changes may drastically alter protein structure and function, as demonstrated by Aharoni et al. [24] and Howard et al. [25] in studies of novel enzyme promiscuity, and by Kaltenbach et al. [26] in the generation of isomerase function in non-isomerase ancestral proteins, among many others [e.g. Refs. [27,28]]. These observations reveal that proteins with highly similar sequences do not necessarily perform the same function.

Beyond these considerations, the vast bioinformatic picture that we have access to through our study of extant organisms is far from complete. Most organisms that have ever lived in the history of the Earth are extinct [29]; except for very few instances of relatively young and exceptionally preserved specimens, this DNA has not been conserved and is not available for study. Furthermore, the fossil-derived extinction statistics for eukaryotic organisms that define the Earth's recent geologic past do not extend to prokaryotes [30,31]. Bacteria and Archaea exhibit much higher rates of turnover at the population level and are thus far more mutable in their exploration of sequence space in the struggle for survival [31–33]. Whereas eukaryotic sequences might be expected to mutate slower over longer time scales because of more restricted fitness topologies and metabolic constraints, prokaryotes yield a less rigid framework for inference, such that extrapolations into the deep past are relatively limited. Finally, the number of genetically sequenced organisms represents only a miniscule fraction of all extant organisms. Recent estimates of extant global biodiversity reach 1 to 6 billion species [34]; however, only ~15,000 completely sequenced genomes are currently available (https://gold.jgi.doe.gov/, accessed September 2018 [35]), and even fewer of these have been fully annotated.

This is a first order summary of the limitations of the available genetic dataset at our disposal. These limitations are non-trivial and fail to provide support for a generalized uniformitarian assumption that extant biomolecules amenable to analysis (even highly essential ones) have persisted, in form or function, throughout Earth's deep history. Thus, uniformitarian assumptions must be developed and critically evaluated on a sequence-by-sequence basis. To address these limitations, recent efforts have sought to reconstruct the sequences of proteins that may have driven key biogeochemical or macroevolutionary changes in Earth's distant past – a prerequisite to assessing whether the resulting phenotypes recapitulate ancestral biogeochemical activity.

3. What is ancestral sequence reconstruction?

By integrating knowledge of the evolutionary process into hypotheses of how ancient proteins may have functioned, it may be possible to recognize the limitations of a uniformitarian approach to biomolecular systems, and to constructively leverage the abundance of available genetic sequence data for questions regarding deep time paleobiology. Ancestral sequence reconstruction, otherwise known as paleogenetics, is a method by which the genetic information of extant organisms is used to computationally estimate the DNA or protein sequences of their ancestors, given a statistical model describing the evolutionary process [reviewed by Refs. [36–39]]. In studies of ancient proteins, ancestral amino acid sequences can be synthesized and experimentally characterized for their functional and biochemical properties. Ancestral sequence reconstruction accompanied by laboratory “resurrections” of ancient proteins can thus be used to answer questions both relating to general mechanisms of protein evolution, adaptation, and the environmental context in which such proteins existed and diversified.

Long before resurrected protein synthesis would become feasible, an incarnation of ancestral sequence reconstruction was first conceived by Linus Pauling and Emile Zuckerkandl [40]. The first experimental applications of ancestral sequence reconstruction were implemented nearly concurrently by the Benner and Wilson groups in the early 1990s [41,42]. Since then, ancestral sequence reconstruction has permitted experimental investigations of the molecular origins of human ethanol metabolism [43], diversification of steroid receptor ligand-binding interactions [44], and dim-light adaptation in early archosaur rhodopsins [45], to name a select few. Recent studies have further extended these approaches by inserting ancient genes in bacterial chromosomes, followed by experimental evolution to study the role of evolutionary contingency [46], as well as by generating transgenic animals with ancient genes to understand the evolutionary trajectory of ethanol tolerance in fruit flies [47].

4. Biogeochemical applications of ancestral sequence reconstruction

Ancestral sequence reconstruction has increasingly been applied to exploring key biological and environmental elements of Earth's past. One reason is that proteins are the primary intermediaries between metabolism and biologically reworked materials in the surrounding environment. Biogeochemical systems would not be possible without the existence and direct involvement of key metabolic proteins. Furthermore, many of these core metabolic proteins are also essential for cellular viability and have therefore been sufficiently conserved to enable the implementation of tractable paleogenetic studies.

Ancestral sequence reconstruction has also become desirable for deep time biogeochemical studies because, despite the fundamental limitations of the bioinformatics dataset, the geologic dataset available

for Earth's past is similarly incomplete. Evidence of ancient environmental conditions is recorded in the structure and distribution of rock formations as well as in the chemical and isotopic compositions of minerals and preserved organic remains. However, continual geologic recycling has imposed a preservational bias on the available rock record. First, deposits of greater age are correspondingly less abundantly exposed [48]. Moreover, particularly ancient sedimentary rocks such as those formed during the early Precambrian reflect only a narrow window into past environments, weighted toward those depositional settings (e.g., nearshore marine) most capable of accumulating and lithifying sediments. Therefore, extrapolating a model of the global ancient biogeochemical environment solely from the information supplied by limited geologic exposures is nonideal. Studies that employ ancestral sequence reconstruction are then conducted with the expectation that this methodology can yield broad, macroevolutionary-scale trends that are not interpretable from the bioinformatic and geological datasets individually [49].

5. Phylogenetic and statistical uncertainty can impact robustness of reconstructed ancient protein function

Computational reconstruction of ancient proteins and experimental characterization of ancient protein function generally proceeds as a series of six steps (Fig. 1). First, a collection of amino acid sequences, inferred to be homologous and capable of capturing the scientific question of interest (such as a hypothesized historical change or gain of function, binding specificity, etc.) are obtained and then aligned along with a group of closely related outgroup sequences. Next, a phylogenetic tree is constructed describing the historical evolutionary relationships between the contemporary molecular descendants and their common ancestors. Ancestral protein sequences at the node or nodes of interest are then typically reconstructed by maximum likelihood (incorporating evolutionary information provided by a user-specified alignment, tree, and substitutional model) or Bayesian inference (which instead simultaneously estimates phylogenetic and substitutional model parameters). Finally, the ancestral proteins are expressed from their synthesized encoding genes, purified (or incorporated into the genome of an appropriate extant host organism; see section entitled "*Linking paleophenotype to biogeochemical data*" below for discussion [46,47]), and characterized for functionality and biochemical properties.

Despite the compelling advantages of using ancestral sequence reconstruction to understand ancient biogeochemical problems, we must also consider the fundamental limits of a paleogenetic approach. Ancestral sequence reconstruction does not imply that a true ancestral protein sequence (or genotype) can be resurrected; rather, we can view an estimated sequence as a consensus of the sequence diversity of a protein within an ancestral population [e.g. Ref. [50]]. With the assumption that a protein exhibited a homogeneous phenotype within an ancient population, the calculated ancestral sequence then can be considered a hypothesis to be tested through laboratory resurrection and functional characterization. Pending the success of these efforts, the resulting phenotype may then be assessed as to whether it may reasonably be tied to the independent geological record.

However, bias resulting from inappropriate methodology applied both to phylogenetic reconstruction (i.e., sequence collection, multiple sequence alignment, and tree reconstruction) and ancestral sequence inference (i.e., selection of substitutional model and statistical threshold) may impact the robustness of resurrected protein function. Validation of computational reconstruction methods has been limited to simulated and experimentally controlled evolutionary histories [51–55] that are unable to capture complicated evolutionary patterns on the timescales desired for biogeochemical applications; it is unclear to what extent similar validation may be possible for macroevolutionary processes that span Earth's deep history. Therefore, utmost care must be taken when designing and implementing paleogenetic studies such that biases do not propagate toward erroneous inferences of ancient biogeochemical systems.

Table 1 summarizes several studies over the last decade that have implemented ancestral sequence reconstruction and have also accounted in some way for ambiguity in inferred ancestral protein function. Such studies demonstrate that sources of uncertainty must be considered and accounted for at each step of the reconstruction process. Below we discuss potential sources of bias and uncertainty associated with each step of the reconstruction process, and we review previous ancestral protein thermostability studies that demonstrate the necessity for caution when implementing ancestral sequence reconstruction for ancient biogeochemical investigations.

5.1. Adequate sampling of homologous protein sequences

In the selection of a protein for a biogeochemically-relevant paleogenetic study, several factors must be considered. First, the collection of contemporary homologous protein sequences must capture sufficient modern sequence diversity for the inference of a targeted ancient phenotype. In addition, the protein sequences themselves must carry enough phylogenetic signal to resolve evolutionary relationships. To resolve particularly deep evolutionary histories, selected protein sequences must be sufficiently conserved so as not to have undergone a great deal of sequence convergence and reversal that would obfuscate inferences of shared ancestry. It is also for this reason that amino acid sequences are typically used for ancestral sequence reconstruction rather than mutable nucleotide sequences, though both can be simultaneously investigated to determine reconstruction robustness to different types of molecular data [e.g. Refs. [45,54,57,63,68]]. These considerations must be taken relative to the antiquity of the biogeochemical problem being investigated.

Several biases associated with the tree reconstruction process can be mitigated by ensuring adequate taxonomic sampling within the protein sequence dataset [e.g. Refs. [69–71]]. Sparse taxonomic sampling may produce phylogenetic trees with several long branches that may incorrectly cluster together due to long branch attraction [reviewed in Ref. [72]]. Selection of an inappropriate, distantly related outgroup will also produce long branches that may affect topology of the ingroup and possibly affect tree rooting [69,73]. The inclusion of more taxa in both the ingroup and outgroup serves to bisect long branches, thereby mitigating erroneous inferences of tree topology, as well as reducing the amount of unobserved sequence changes needed to be inferred along branches [74,75]. In addition, greater taxonomic representation improves the estimation of evolutionary model parameters for maximum likelihood and Bayesian analyses [76] Finally, adequate taxonomic sampling among basal lineages relative to the ancestral node of interest is essential. Insufficient representation among these lineages may shift the target of investigation to a more recently diverged protein ancestor–one not necessarily relevant to the desired scientific question.

5.2. Avoiding homology violations in multiple sequence alignments

An amino acid alignment implies site homology among the aligned sequences; violations of this assumption can lead to erroneous estimations of phylogenetic relationships and ancestral sequences downstream within the paleogenetic study. Highly diverged sequences, and/or sequences containing a high proportion of insertions and deletions, are especially prone to misalignment. For simulated protein sequences, Vialle et al. [55] showed that MAFFT [77] and PRANK [78] alignment algorithms performed better than others in difficult-to-align datasets in the accuracy of downstream reconstructed ancestral sequences. In particular, the PRANK algorithm, which incorporates phylogenetic information into the alignment procedure, was able to better estimate insertions and deletions. Despite such advantages, Anisimova et al. [79] found that these sophisticated methods have been generally underutilized, and specifically cautioned against manual alignment adjustments, as doing so may impede reproducibility of the scientific

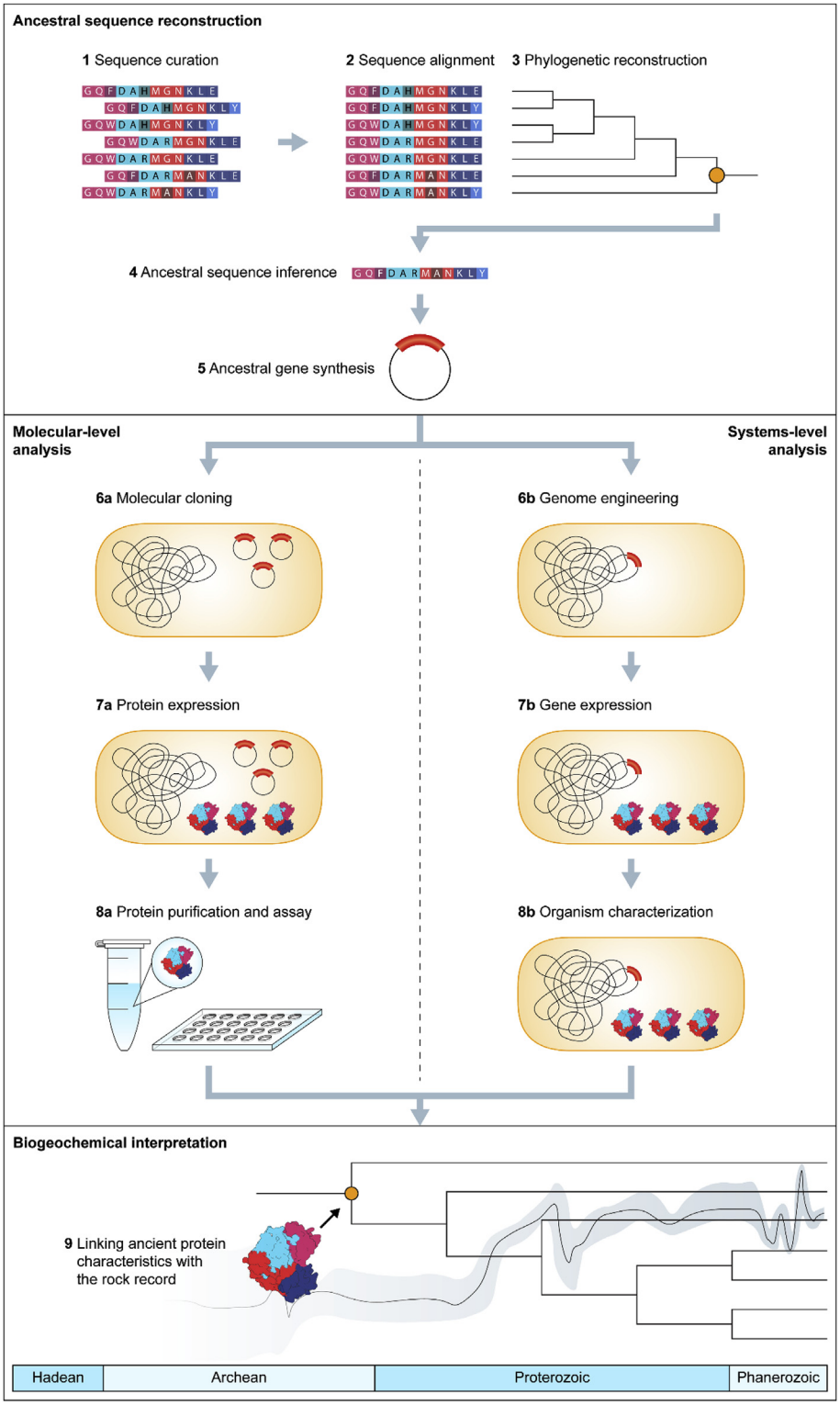

Fig. 1. Proposed strategy for applying ancestral sequence reconstruction to biogeochemical studies. This diagram outlines the steps required to reconstruct and characterize an ancient protein to be mapped against the biogeochemical record. (1) Curate a dataset of extant protein sequences. (2) Perform a multiple sequence alignment of extant sequences. (3) Reconstruct a phylogenetic tree based on the alignment in accordance with a best-fitting evolutionary model. (4) Infer the ancestral sequence at the phylogenetic node of interest with maximum likelihood or Bayesian algorithms. (5) Synthesize the gene encoding region of the inferred ancestral protein sequence into a plasmid for molecular- or systems-level (*in vivo*) experimental analysis. (6a) Transform the encoding ancestral gene sequence into an appropriate expression system or (6b) modify an appropriate host organism genome to incorporate the ancestral gene. (7a) Express the ancestral protein from the incorporated plasmid or (7b) modified genome. (8a) Purify the expressed protein for biochemical characterization or (8b) assess the reconstructed paleophenotype at the organismal scale *in vivo*. Protein purification and characterization methods must be tailored to the protein system of interest and tested with more recent ancestors to ensure viability. (9) Integrate results from both molecular- and systems-level experimental analyses with aspects of the biogeochemical record thought to have been modulated by the ancient protein.

methodology. The appropriate selection of alignment algorithm ultimately depends on the nature of the protein dataset. For a detailed review of available methods, including those that account for phylogenetic and structural information as well as the identification of unreliably aligned sequence regions, see Anisimova et al. [79].

5.3. Impact of phylogenetic tree reconstruction methods on ancestral inference

Phylogenetic tree reconstruction methods use a multiple sequence alignment as input and generate a tree topology and branch lengths describing the vertical evolutionary process. State-of-the-art methods typically used in paleogenetic studies operate within either a maximum likelihood (e.g., PhyML [80], GARLI [81], IQ-TREE [82], and RaxML [83]) or Bayesian framework (e.g., MrBayes [84], PhyloBayes [85], and

Table 1
A selection of ancestral sequence reconstruction studies published within the last 10 years and relevant to biogeochemical applications. Check marks indicate studies that have explicitly tested more than one method of modern sequence sampling, alignment, tree reconstruction, and/or ancestral sequence inference (including different substitutional models, inference algorithms, and inference of near-ancestors) in accounting for reconstruction uncertainty. This table does not represent an exhaustive listing of strategies for mitigating bias; therefore, other unrepresented but potentially effective measures (as discussed in text) may also have been implemented in these investigations. *P. Shih, personal communication.*

Year	Reference	Reconstructed protein	Tested multiple methods of:			
			Modern sequence sampling	Alignment	Tree reconstruction	Ancestral sequence inference
2008	[50]	elongation factor Tu			✓	✓
2011	[56]	thioredoxin				
2012	[57]	3-isopropylmalate dehydrogenase				✓
2013	[65]	nucleoside diphosphate kinase	✓		✓	✓
2013	[66]	class-A β-lactamase				✓
2014	[59]	ribonuclease H1				✓
2015	[60]	nucleoside diphosphate kinase	✓		✓	✓
2016	[12]	RuBisCO			✓*	
2017	[64]	nucleoside diphosphate kinase			✓	
2017	[13]	RuBisCO		✓	✓	✓
2017	[14]	carbonic anhydrase		✓	✓	✓
2017	[67]	adenylate kinase				✓

BAli-Phy [86]). Integrated into many of these methods are means of assessing phylogenetic confidence, including bootstrap proportion [87] and branch support evaluation [reviewed in Ref. [88]]. Substitutional model fits can additionally be ranked by algorithms implemented in ProtTest [89].

The degree to which ancestral sequence inference and resulting experimental phenotypic characterization is modulated by phylogenetic reconstruction methods has only been investigated to a limited extent. One relatively simple strategy for capturing this uncertainty is to simultaneously infer multiple phylogenetic trees to describe the protein sequence dataset and carry out the remainder of the paleogenetic study using each [e.g. Refs. [13,50,64,65]]. In addition, well-regarded species trees from the literature can instead be used [45,50,58,61,90]. Groussin et al. suggested that species-tree-aware phylogenetic methods, which incorporate models accounting for genome-scale evolution, can increase accuracy of ancestral sequence reconstruction. Differences between maximum likelihood and Bayesian phylogenetic inference have additionally been evaluated by Hanson-Smith et al. [53], who concluded that, though Bayesian methods may intuitively seem preferable because they integrate phylogenetic uncertainty into ancestral sequence inference, they did not necessarily reduce ancestral sequence ambiguity.

5.4. Accommodating statistical uncertainty in ancestral sequence inference

Methods for ancestral sequence inference can also be implemented using maximum likelihood, e.g., PAML [91], and Bayesian approaches. Maximum likelihood methods incorporate user-specified phylogenetic trees and substitution models to describe the protein dataset, whereas fully Bayesian methods integrate over the uncertainty in these phylogenetic parameters. These different frameworks for reconstructing protein ancestors can be simultaneously investigated [e.g. Refs. [54,57,60]] along with different accompanying substitution models [e.g. Refs. [13,26,63,90]]. In addition, these algorithms can be applied to insertion and deletion reconstruction to infer gap placement within the inferred ancestral sequence [92,93].

Some practitioners of paleogenetic studies have in the past used single point estimates of the most likely protein ancestor from each reconstruction – that is to say, the ancestral sequence having the greatest probability of producing the modern sequence dataset given the evolutionary model. Though this can be preferable due to both computational and downstream experimental limitations, the problem with this approach can be considered as follows: Consider a 200-amino-acid protein sequence. Even if each amino acid is estimated with 0.90 probability, the probability of the sequence itself is exceedingly low, only $0.90^{200} \approx 1 \times 10^{-10}$. Therefore, there is a space of comparably plausible near-ancestral sequences within which ancestral protein function would ideally be experimentally characterized.

Reconstructions of inferred near-ancestors along with the “most likely” ancestor, deal both with statistical uncertainty associated with different probabilistic frameworks, as well as the oftentimes arbitrary nature of selected cut-offs for ancestral sequence confidence. Some studies have incorporated this approach either by introducing site-by-site “second-best” alternate residues to ancestral variants [e.g., Refs. [68,94–96]], or by introducing these alternate residues simultaneously in a single sequence [e.g. Refs. [62,65,97]]. A full statistical integration of sequence uncertainty may also be made by randomly sampling ancestors from the posterior probability distribution of state reconstructions within a Bayesian framework [36,52]. The resulting phenotypic interpretation of the “true ancestor” would average from the individual stabilities of each “near-ancestor’ weighted by their inferred sequence posterior probabilities [e.g. Refs. [50,98]]. Eick et al. [99] have investigated these different approaches of generating and experimentally characterizing near-ancestors, finding that qualitative functionality (e.g., ligand specificity) of inferred ancestors of a variety of protein families is typically robust to the method of capturing sequence uncertainty. However, they did find that Bayesian sampling was often biased toward non-functioning protein reconstructions, possibly because such sampling is not limited to higher probability residues. Finally, though qualitative protein function did remain essentially the same across statistical uncertainty of the ancestors, quantitative characterization (e.g., kinetic parameters) did show significant variability. Since enzymatic kinetic parameters impact isotopic fractionation of molecular substrates [100–104], accounting for the distribution of ancestral protein functionality across statistical uncertainty is likely a nontrivial necessity for many biogeochemical paleogenetic studies.

Does the most probable ancestor represent the true ancestral phenotype? Development of high-throughput gene syntheses and subsequent enzyme functionality assays have enabled experimental exploration of ancestral sequence probability space to address this question. Bar-Rogovsky et al. [98] developed a framework to identify ambiguously reconstructed residues that were likely to impact protein function and conformation, and subsequently constructed libraries to experimentally characterize the weighted sequence variants. Functional screening of hundreds of ancestral paraoxonase variants allowed the authors to statistically describe the phenotype variation within this enzyme pool. Ancestral paraoxonase libraries demonstrated that reconstructed enzyme variants exhibited variation in catalysis, even though the phylogenetic reconstruction output suggested that all the sequences comprising these libraries could be considered statistically

plausible. The phenotype of the most probable paraoxonase ancestors then could not be interpreted as a good representative of the "true" ancestral phenotype [98]. This combinatorial study addressed a valid concern regarding inferred ancestral sequence ambiguity, by directly connecting reconstructed sequence uncertainty with protein functional variability.

In summary, uncertainty in ancestral sequence inference can be derived from biases associated with essentially each step of the phylogenetic reconstruction process. The potential for compounding error through the entire process to experimental characterization necessitates deliberate evaluation of phylogenetic procedures. However, due to the inherent complexity of molecular evolution, significant phylogenetic uncertainty cannot always be avoided. Therefore, the sensitivity of phenotypic variability of reconstructed proteins to sequence sampling, alignment, tree reconstruction, and ancestral sequence inference methodologies should be carefully investigated. Below we summarize some studies that are of interest to the biogeochemical community that addressed such uncertainties.

5.5. Ancient protein thermostability – a case study in paleogenetic uncertainty

Over the last decade, various studies have applied ancestral sequence reconstruction and subsequent experimental characterization to infer ancient protein thermostabilities [50,56,57,59,64–67,105]. Some have additionally presumed past trends in global Earth temperatures by the thermostabilities of a diversity of ancestral proteins including elongation factors [50], thioredoxins [56], β-lactamases [66], and kinases [64,67]. The use of paleogenetics to probe Earth temperature histories is desirable owing to uncertainties associated with Precambrian geochemical temperature proxies [106–109]. Ancestral sequence reconstruction then provides an independent means by which to assess interpretations of geochemical proxies and fill a knowledge gap inherent to an incomplete geologic record.

The majority of these studies have found that ancient protein thermostabilities exceed those of their extant descendants, with stability generally increasing with estimated divergence age. These findings have been found to be robust to tested sources of phylogenetic uncertainty, including effects of taxonomic sampling [60,65], phylogenetic tree topology [50,64,65,105], substitutional model [50,105], and maximum likelihood versus Bayesian inference [57,67,105].

However, Williams et al. [52] suggested that a systematic "consensus bias" toward common amino acids at an alignment position may cause the inference of inaccurate and overly stable ancestral proteins. Introduction of consensus residues are in fact often found to stabilize proteins [110,111]. Still, experimental characterizations of consensus proteins find that they are sometimes inactive or not expressed, and do not necessarily display greater thermostabilities than inferred ancestors [112–114]. It is therefore apparent that properties of ancestral proteins do not always simply mimic those of the consensus [112], but it is still unclear to what extent consensus bias may affect ancestral phenotype and whether that effect can be systematically quantified [115,116]. It is possible that minimization of consensus bias effects may be achieved by expanding ancestral sequence inference to the suite of statistically plausible near-ancestors, as described in the previous section entitled "*Accommodating statistical uncertainty in ancestral sequence inference*" [50,52,65,96,115].

Investigations of protein thermostabilities represent only one way to connect paleogenetic approaches to the ancient Earth environment. The identification of sources of systematic bias in these studies, as well as potentially useful mitigation tactics, demonstrate that future paleogenetic studies must proceed cautiously.

5.6. Linking resurrected paleophenotypes to biogeochemical data

As we have outlined in the previous section, experimental applications of ancestral sequence reconstruction must ensure that phenotypic inferences are not artifacts arising from phylogenetic and/or statistical bias. However, to connect ancient protein function to past biogeochemical cycling, other criteria must also be met.

Biosignatures identified in the geologic record, namely isotopic compositions of preserved minerals and organic matter, were not produced by ancient proteins in isolation but were rather the expression of such proteins within a complex organismal system. As such, phenotypic inferences made from single-protein resurrections may not be fully applicable to comparisons with biosignatures found in the geologic record.

By incorporating the encoding genes of ancestral proteins into modified microbial genomes, it is possible that the ancestral phenotypes that would have produced ancient biosignatures can be more faithfully reproduced. We have previously proposed a set of criteria for ancestral "paleophenotype" inference, or the reconstruction of ancient biosignatures within entire functioning molecular systems [49] (Fig. 1). In summary, paleophenotypic reconstruction can be feasible and yield insights not possible by single-protein reconstruction if (1) the biosignature is geochronologically constrained and correlated with enzymatic function and/or phylogenetic divergence (i.e., the geology is sufficiently detailed; though not within the scope of this article, geochronology can potentially be complemented by well-constrained molecular clock analyses of the target phylogeny), (2) the geochemical output of a protein phenotype can be well-characterized and compared to geologically preserved biosignatures (i.e., the phenotype is sufficiently constrained), and (3) the modified host organism expressing the ancestral protein is similar enough to the ancestral host organism and well-studied as a model organism to reliably reproduce ancient phenotypic characters (i.e., the extant analog system sufficiently reflects ancestral physiology and ecology).

Paleophenotypic reconstruction may overcome ambiguities between *in vitro* and *in vivo* protein function and offer a strategy to mitigate erroneous phenotypic inference due to phylogenetic or statistical reconstruction bias. Just as resurrected enzyme functionality is oftentimes taken as evidence for the reliability of the reconstruction, an active resurrected enzyme able to participate within an organismal system can confirm the internal consistency of a reconstruction. This strategy additionally provides a means in the future toward reconstructing entire functional ancestral metabolic networks, which together may recapitulate the production of biosignatures comparable to those identified in the geologic record.

5.7. Paleophenotype reconstruction across the Great Oxidation Event

Considering the need to prudently address sources of bias inherent to paleophenotypic reconstruction, it remains to be seen how the uncertainties of ancestral sequence reconstruction will impact inferred protein phenotypes associated with ancient biogeochemical cycling. One path forward is to consider the application of paleogenetics toward ancient proteins coincident with the Great Oxidation Event [117], arguably one of the most significant biogeochemical transitions in Earth's history.

Earliest microfossil evidence of cyanobacteria dating to approximately 2.0–2.5 Ga suggests that the evolution of oxygenic photosynthesis largely overlapped with and likely drove the oxygenation of the Earth environment during the Great Oxidation Event [118–121]. The rise of photosynthetically produced oxygen accompanying the proliferation of cyanobacteria would have imposed new selective pressures on anaerobic prokaryotes that had previously dominated the anoxic Archean surface environment. Thus, with changes in microbial diversity, as well as the global metabolome, came the compositional shift of gaseous species in the atmosphere and oceans of the planet. A reasonable null hypothesis is that essential enzymes present in crucial cellular networks remained largely unimpacted due to internal evolutionary constraints [122–124]. However, in light of the ready reactivity

of molecular oxygen, it is also possible that enzymes may have undergone evolutionary adaptations to alter kinetics and substrate specificities in response to atmospheric changes [12,13].

Thus, with changes in microbial diversity, as well as the global metabolome, came the compositional shift of gaseous species in the atmosphere and oceans of the planet. It is tempting to propose that enzymes may have undergone evolutionary adaptations to alter kinetics and substrate specificities in response to these gaseous changes. The molecular systems of ancient organisms may have then adapted to newly imposed oxidative stress during the Great Oxidation Event [8–11]. This opens a geochronologically and paleobiologically constrained window of investigation into the mechanisms by which life may have adapted to this significant biogeochemical transition.

5.8. Reconstructed ancestral RuBisCO enzymes

RuBisCO (ribulose 1,5-bisphosphate carboxylase/oxygenase) represents a conspicuous enzyme target for paleophenotype resurrection relevant to the Great Oxidation Event. Though some investigations have indicated that the Wood-Ljungdahl pathway may be more ancient [125,126], Form I RuBisCO, which mediates carbon fixation in oxygenic photosynthesizing cyanobacteria and land plants, is thought to have been the primary enzymatic facilitator of carbon sequestration into biological and geological reservoirs through much of Earth's history [127,128]. Its association with oxygenic photosynthesis allows its evolutionary history to be mapped to the first whiffs of oxygen more than two billion years ago [129]. The RuBisCO family also encompasses other forms, some of which participate in alternate carbon fixation pathways in anaerobes or possess more primitive non-carboxylase/oxygenase function. In sum, the evolution of RuBisCO enzymes has traversed and been intimately exposed to the changing redox conditions of the Earth environment over billions of years.

RuBisCO also possesses a functional quirk by having both carboxylase and oxygenase activity, significantly reducing the efficiency of carbon fixation. It is possible that the modern CO_2/O_2 specificity of RuBisCO is a primitive relict of ancestral optimization under an ancient CO_2-enriched atmosphere, when low CO_2 specificity would not have imposed a significant fitness cost before the accumulation of atmospheric oxygen. Evidence suggesting a tight linkage between CO_2 specificity and carbon isotopic fractionation of RuBisCO carboxylase activity [102] suggests, importantly, that primitive characteristics of ancient RuBisCO may be calibrated to the isotopic compositions of carbon preserved in sedimentary rocks.

RuBisCO enzymes then satisfy all criteria for paleophenotypic study. First, ancient RuBisCO activity can be linked to geochronologically constrained biosignatures available in the rock record. Second, incorporation of RuBisCO genes into microbial genomes may be possible by use of model organisms (i.e., cyanobacteria), whose ancestors exhibited comparable morphology and habitat [130,131]. Third, the substrate specificity and isotopic fractionation of resurrected RuBisCO enzymes may serve as proxies for the interpretation of the rock record [132].[1]

Recently, certain studies have endeavored to infer ancestral sequences of RuBisCO and connect aspects of sequence composition and kinetic parameters of laboratory-resurrected enzymes to adaptive shifts potentially associated with the Great Oxidation Event [12,13]. Paleogenetic characterizations conducted by Shih et al. [12] suggest that RuBisCO underwent different historical optimization strategies (CO_2 specificity versus CO_2 turnover) in eukaryotic and cyanobacterial lineages with the increase in atmospheric oxygen through the Proterozoic. However, a relative lack of broad taxonomic sampling and narrow outgroup selection in the phylogenetic reconstruction used for ancestral sequence inference may have impacted the reliability of the kinetic characterizations, as well as the relative divergence time calibrations of the resurrected nodes in relation to the Great Oxidation Event [13]. Future studies following in this line of investigation must consider similar possible pitfalls, as well as potentially fruitful strategies to accommodate ancestral sequence uncertainty (i.e., reconstruction model comparisons, near-ancestor sampling, etc. as discussed in previous sections). Nevertheless, these initial explorations of ancient RuBisCO evolution provide a starting point for future, integrative paleophenotypic reconstruction.

5.9. Reconstructed ancient nitrogenase enzymes

Nitrogenase (EC 1.18.6.1) is an enzyme uniquely capable of catalyzing the reduction of inert atmospheric dinitrogen to ammonia. Though nitrogenase is possessed by a broad diversity of both extant anaerobic and aerobic prokaryotes, it is irreversibly inhibited by oxygen [e.g. Ref. [135]]. Therefore, aerobic organisms today exhibit strategies to shield nitrogen-fixation activity from the deleterious effects of oxygen, including both temporal and spatial separation of nitrogen-fixing and oxygen-producing cellular reactions, as well as active oxygen consumption to maintain sufficiently minimal levels within the cell [3,136].

The most abundant modern form of nitrogenase binds a molybdenum ion at the active site, though alternative, albeit less catalytically efficient nitrogenases that each bind either a vanadium or an iron ion are also known. These alternative nitrogenases are only expressed under anoxic, molybdenum-limiting conditions by organisms that typically otherwise express the canonical molybdenum nitrogenase [137]. Because geochemical evidence suggests that molybdenum was likely scarce in Earth's oceans until after the onset of oxidative weathering accompanying the Great Oxidation Event [2], this alternative nitrogenase behavior suggests a potentially primitive sensitivity to the ambient redox environment [7].

The apparent connection between records of trace metal availability and nitrogenase metal-binding affinities suggest that alternative vanadium and iron nitrogenases might have been ancestral to molybdenum nitrogenases, and potentially present in the last universal common ancestor [7,138]. However, though multiple phylogenetic reconstructions of nitrogenase have supported a shared evolutionary origin [138–140], the precise evolutionary trajectory and timing of origin in relation to the Great Oxidation Event inferred by phylogenetic analyses do not conform to geochemical interpretations. Boyd et al. [139,140] have instead suggested that molybdenum nitrogenase was ancestral, and potentially did not evolve until after the rise of oxygenation ~1.5–2.2 Ga.

Predictions generated by phylogenetic studies regarding the ancestral metal-binding states of nitrogenases may be tested by the experimental characterization of resurrected nitrogenase in the context of ancient geochemical evidence. Modern nitrogenases exhibit isotopic fractionation variability dependent on dinitrogen reduction by different metal-binding enzyme forms [104,141] and ambient redox-sensitive trace metal availability [142,143]. These data suggest that well-characterized isotopic fractionation behavior of resurrected nitrogenases may permit connection to the geochemical isotope record of nitrogen fixation across ancient redox transitions [144–146]. The reconstruction of active, ancestral nitrogenase is likely to be a significant challenge given the complex gene cluster and associated proteins necessary for cofactor biosynthesis present in modern diazotrophs [139,147]. The impact of these co-evolved proteins on the experimental properties of ancestral nitrogenase have yet to be examined.

Nevertheless, the combination of a geochronologically constrained nitrogen isotopic record and the potential to experimentally characterize isotope fractionation behavior and metal-binding of nitrogenase opens a window for the use of paleophenotypic

[1] RuBisCO averages a carbon isotope fractionation value of approximately −25‰, corresponding closely to the average fractionation values observed in the geologic record [133,134].

reconstruction. Sequence and structural evolution of ancestral nitrogenase may reveal signatures of phenotypic shifts in response to a changing ancient geochemical environment [148]. Moreover, by incorporating ancestral nitrogenase genes into the genomes of extant nitrogen-fixing microbial hosts, the behavior of ancient nitrogenase within a complex cellular environment can be assessed. Because the isotopic discrimination behavior of ancient nitrogenases would be reflected in the biomass of engineered modern host organisms, reliable analogs to isotopic signatures in preserved organic matter can be recapitulated and connected with metal-binding affinity. Thus, the evolutionary trajectory of nitrogenase through progressive environmental oxygenation might be revealed by the integration of phylogenetics and experimental functional studies of resurrected proteins.

6. Conclusions

The application of paleogenetic studies to problems of ancient biogeochemical cycling is in its infancy. Furthermore, paleogenetics has not yet been applied sufficiently to warrant a generalized uniformitarian assumption of deep time (i.e., Precambrian) macroevolutionary functional conservation for resurrected ancestral proteins. However, with careful experimental design and a detailed accounting of methodological biases associated with ancestral sequence reconstruction tools, resurrected proteins may recapitulate some key functional attributes of ancient enzymes that have shaped biogeochemical cycling. RuBisCO and nitrogenase represent just two examples of enzymes with which paleophenotypic reconstructions might resolve molecular adaptive responses of the Great Oxidation Event, and their consequential effects on planet-scale biogeochemical cycling records. Looking forward, paleophenotypic, *in vivo* resurrections of molecular systems have the potential to provide insights for understanding ancient biogeochemistry, and for future integrations of geologic and genomic historical records of life.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.freeradbiomed.2019.03.033.

Funding

This work was supported by the National Science Foundation (BK #1724090); by John Templeton Foundation (BK, #58562 and #61239); the NASA Exobiology and Evolutionary Program (BK, #H006201406); and the NASA Astrobiology Institute Postdoctoral Program (AKG).

References

[1] H.D. Holland, Volcanic gases, black smokers, and the great oxidation event, Geochem. Cosmochim. Acta 66 (21) (2002) 3811–3826.
[2] T.W. Lyons, C.T. Reinhard, N.J. Planavsky, The rise of oxygen in Earth's early ocean and atmosphere, Nature 506 (7488) (2014) 307–315.
[3] I. Berman-Frank, P. Lundgren, P. Falkowski, Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria, Res. Microbiol. 154 (3) (2003) 157–164.
[4] A.H. Knoll, The geological consequences of evolution, Geobiology 1 (1) (2003) 3–14.
[5] J.A. Imlay, The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium, Nat. Rev. Microbiol. 11 (7) (2013) 443–454.
[6] L. Margulis, D. Sagan, Microcosmos: Four Billion Years of Evolution from Our Microbial Ancestors, 1st Touchstone ed., Simon & Schuster, New York, 1991.
[7] A.D. Anbar, Proterozoic ocean chemistry and evolution: a bioinorganic bridge? Science 297 (5584) (2002) 1137–1142.
[8] J. Raymond, D. Segrè, The effect of oxygen on biochemical networks and the evolution of complex life, Science 311 (5768) (2006) 1764–1767.
[9] P.G. Falkowski, L.V. Godfrey, Electrons, life and the evolution of Earth's oxygen cycle, Phil. Trans. Biol. Sci. 363 (1504) (2008) 2705–2716.
[10] J.W. Schopf, Geological evidence of oxygenic photosynthesis and the biotic response to the 2400-2200 Ma “Great Oxidation Event”, Biochemistry (Mosc.) 79 (3) (2014) 165–177.
[11] E.K. Moore, B.I. Jelen, D. Giovannelli, H. Raanan, P.G. Falkowski, Metal availability and the expanding network of microbial metabolisms in the Archaean eon, Nat. Geosci. 10 (9) (2017) 629–636.
[12] P.M. Shih, A. Occhialini, J.C. Cameron, P.J. Andralojc, M.A.J. Parry, C.A. Kerfeld, Biochemical characterization of predicted precambrian RuBisCO, Nat. Commun. 7 (2016) 10382.
[13] B. Kacar, V. Hanson-Smith, Z.R. Adam, N. Boekelheide, Constraining the timing of the great oxidation event within the rubisco phylogenetic tree, Geobiology 15 (5) (2017) 628–640.
[14] B. Kacar, L. Guy, E. Smith, J. Baross, Resurrecting ancestral genes in bacteria to interpret ancient biosignatures, Phil. Trans. Math. Phys. Eng. Sci. 375 (2109) (2017) 20160352.
[15] E.V. Koonin, Orthologs, paralogs, and evolutionary genomics, Annu. Rev. Genet. 39 (2005) 309–338.
[16] M. Kimura, T. Ohta, On some principles governing molecular evolution, Proc. Natl. Acad. Sci. U.S.A. 71 (7) (1974) 2848–2852.
[17] A.C. Wilson, S.S. Carlson, T.J. White, Biochemical evolution, Annu. Rev. Biochem. 46 (1977) 573–639.
[18] G. Karp, Cell and Molecular Biology : Concepts and Experiments, fifth ed., John Wiley & Sons, Chichester, England ; Hoboken, New Jersey, 2008.
[19] J. Zhang, J.-R. Yang, Determinants of the rate of protein sequence evolution, Nat. Rev. Genet. 16 (7) (2015) 409–420.
[20] R.V. Eck, M.O. Dayhoff, Evolution of the structure of ferredoxin based on living relics of primitive amino Acid sequences, Science 152 (3720) (1966) 363–366.
[21] I.K. Jordan, I.B. Rogozin, Y.I. Wolf, E.V. Koonin, Essential genes are more evolutionarily conserved than are nonessential genes in bacteria, Genome Res. 12 (6) (2002) 962–968.
[22] J.A. Capra, M. Singh, Predicting functionally important residues from sequence conservation, Bioinformatics 23 (15) (2007) 1875–1882.
[23] D. Lee, O. Redfern, C. Orengo, Predicting protein function from sequence and structure, Nat. Rev. Mol. Cell Biol. 8 (12) (2007) 995–1005.
[24] A. Aharoni, L. Gaidukov, O. Khersonsky, S.M. Gould, C. Roodveldt, D.S. Tawfik, The 'evolvability' of promiscuous protein functions, Nat. Genet. 37 (1) (2005) 73–76.
[25] C.J. Howard, V. Hanson-Smith, K.J. Kennedy, C.J. Miller, H.J. Lou, A.D. Johnson, B.E. Turk, L.J. Holt, Ancestral resurrection reveals evolutionary mechanisms of kinase plasticity, eLife 3 (2014).
[26] M. Kaltenbach, J.R. Burke, M. Dindo, A. Pabis, F.S. Munsberg, A. Rabin, S.C.L. Kamerlin, J.P. Noel, D.S. Tawfik, Evolution of chalcone isomerase from a noncatalytic ancestor, Nat. Chem. Biol. 14 (6) (2018) 548–555.
[27] I. Matsumura, A.D. Ellington, In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates, J. Mol. Biol. 305 (2) (2001) 331–339.
[28] B.M.A. van Vugt-Lussenburg, M.C. Damsten, D.M. Maasdijk, N.P.E. Vermeulen, J.N.M. Commandeur, Heterotropic and homotropic cooperativity by a drug-metabolising mutant of cytochrome P450 BM3, Biochem. Biophys. Res. Commun. 346 (3) (2006) 810–818.
[29] D. Jablonski, Extinctions in the fossil record, Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 344 (1307) (1994) 11–17.
[30] M. Weinbauer, F. Rassoulzadegan, Extinction of microbes: evidence and potential consequences, Endanger. Species Res. 3 (2007) 205–215.
[31] J. Marin, F.U. Battistuzzi, A.C. Brown, S.B. Hedges, The timetree of prokaryotes: new insights into their evolution and speciation, Mol. Biol. Evol. (2016) msw245.
[32] W.B. Whitman, D.C. Coleman, W.J. Wiebe, Prokaryotes: the unseen majority, Proc. Natl. Acad. Sci. U.S.A. 95 (12) (1998) 6578–6583.
[33] F.M. Cohan, Bacterial species and speciation, Syst. Biol. 50 (4) (2001) 513–524.
[34] B.B. Larsen, E.C. Miller, M.K. Rhodes, J.J. Wiens, Inordinate fondness multiplied and redistributed: the number of species on Earth and the new pie of life, Q. Rev. Biol. 92 (3) (2017) 229–265.
[35] S. Mukherjee, D. Stamatis, J. Bertsch, G. Ovchinnikova, O. Verezemska, M. Isbandi, A.D. Thomas, R. Ali, K. Sharma, N.C. Kyrpides, T.B.K. Reddy, Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements, Nucleic Acids Res. 45 (D1) (2017) D446–D456.
[36] J.W. Thornton, Resurrecting ancient genes: experimental analysis of extinct molecules, Nat. Rev. Genet. 5 (5) (2004) 366–375.
[37] S.A. Benner, D.A. Liberles, The Early Days of Paleogenetics: Connecting Molecules to the Planet, Ancestral Sequence Reconstruction, Oxford University Press, 2007, pp. 3–19.
[38] Y. Gumulya, E.M.J. Gillam, Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro approach to protein engineering, Biochem. J. 474 (1) (2017) 1–19.
[39] R. Merkl, R. Sterner, Ancestral protein reconstruction: techniques and applications, Biol. Chem. 397 (1) (2016) 1–21.
[40] L. Pauling, E. Zuckerkandl, T. Henriksen, R. Lövstad, Chemical paleogenetics. Molecular "restoration studies" of extinct forms of life, Acta Chem. Scand. 17 supl (1963) 9–16.
[41] J. Stackhouse, S.R. Presnell, G.M. McGeehan, K.P. Nambiar, S.A. Benner, The ribonuclease from an extinct bovid ruminant, FEBS (Fed. Eur. Biochem. Soc.) Lett. 262 (1) (1990) 104–106.
[42] B.A. Malcolm, K.P. Wilson, B.W. Matthews, J.F. Kirsch, A.C. Wilson, Ancestral lysozymes reconstructed, neutrality tested and thermostability linked to hydrocarbon packing, Nature 345 (6270) (1990) 86–89.
[43] M.A. Carrigan, O. Uryasev, C.B. Frye, B.L. Eckman, C.R. Myers, T.D. Hurley, S.A. Benner, Hominids adapted to metabolize ethanol long before human-directed fermentation, Proc. Natl. Acad. Sci. U.S.A. 112 (2) (2015) 458–463.
[44] J.W. Thornton, Resurrecting the ancestral steroid receptor: ancient origin of

estrogen signaling, Science 301 (5640) (2003) 1714–1717.
[45] B.S.W. Chang, K. Jönsson, M.A. Kazmi, M.J. Donoghue, T.P. Sakmar, Recreating a functional ancestral archosaur visual pigment, Mol. Biol. Evol. 19 (9) (2002) 1483–1489.
[46] B. Kacar, X. Ge, S. Sanyal, E.A. Gaucher, Experimental evolution of Escherichia coli harboring an ancient translation protein, J. Mol. Evol. 84 (2–3) (2017) 69–84.
[47] M.A. Siddiq, D.W. Loehlin, K.L. Montooth, J.W. Thornton, Experimental test and refutation of a classic case of molecular adaptation in Drosophila melanogaster, Nat. Ecol. Evol. 1 (2) (2017) 0025.
[48] H. Blatt, R.L. Jones, Proportions of exposed igneous, metamorphic, and sedimentary rocks, Geol. Soc. Am. Bull. 86 (8) (1975) 1085.
[49] B. Kacar, L. Guy, E. Smith, J. Baross, Resurrecting ancestral genes in bacteria to interpret ancient biosignatures, Phil. Trans. Math. Phys. Eng. Sci. 375 (2109) (2017) 20160352.
[50] E.A. Gaucher, S. Govindarajan, O.K. Ganesh, Palaeotemperature trend for Precambrian life inferred from resurrected proteins, Nature 451 (7179) (2008) 704–707.
[51] D. Hillis, J. Bull, M. White, M. Badgett, I. Molineux, Experimental phylogenetics: generation of a known phylogeny, Science 255 (5044) (1992) 589–592.
[52] P.D. Williams, D.D. Pollock, B.P. Blackburne, R.A. Goldstein, Assessing the accuracy of ancestral protein reconstruction methods, PLoS Comput. Biol. 2 (6) (2006) e69.
[53] V. Hanson-Smith, B. Kolaczkowski, J.W. Thornton, Robustness of ancestral sequence reconstruction to phylogenetic uncertainty, Mol. Biol. Evol. 27 (9) (2010) 1988–1999.
[54] R.N. Randall, C.E. Radford, K.A. Roof, D.K. Natarajan, E.A. Gaucher, An experimental phylogeny to benchmark ancestral sequence reconstruction, Nat. Commun. 7 (2016) 12847.
[55] R.A. Vialle, A.U. Tamuri, N. Goldman, Alignment modulates ancestral sequence reconstruction accuracy, Mol. Biol. Evol. 35 (7) (2018) 1783–1797.
[56] R. Perez-Jimenez, A. Inglés-Prieto, Z.-M. Zhao, I. Sanchez-Romero, J. Alegre-Cebollada, P. Kosuri, S. Garcia-Manyes, T.J. Kappock, M. Tanokura, A. Holmgren, J.M. Sanchez-Ruiz, E.A. Gaucher, J.M. Fernandez, Single-molecule paleoenzymology probes the chemistry of resurrected enzymes, Nat. Struct. Mol. Biol. 18 (5) (2011) 592–596.
[57] J.K. Hobbs, C. Shepherd, D.J. Saul, N.J. Demetras, S. Haaning, C.R. Monk, R.M. Daniel, V.L. Arcus, On the origin and evolution of thermophily: reconstruction of functional precambrian enzymes from ancestors of Bacillus, Mol. Biol. Evol. 29 (2) (2012) 825–835.
[58] I. van Hazel, A. Sabouhanian, L. Day, J.A. Endler, B.S.W. Chang, Functional characterization of spectral tuning mechanisms in the great bowerbird short-wavelength sensitive visual pigment (SWS1), and the origins of UV/violet vision in passerines and parrots, BMC Evol. Biol. 13 (1) (2013) 250.
[59] K.M. Hart, M.J. Harms, B.H. Schmidt, C. Elya, J.W. Thornton, S. Marqusee, Thermodynamic system drift in protein evolution, PLoS Biol. 12 (11) (2014) e1001994.
[60] S. Akanuma, S.-i. Yokobori, Y. Nakajima, M. Bessho, A. Yamagishi, Robustness of predictions of extremely thermally stable proteins in ancient organisms, Evolution 69 (11) (2015) 2954–2962.
[61] C. Bickelmann, J.M. Morrow, J. Du, R.K. Schott, I. van Hazel, S. Lim, J. Müller, B.S.W. Chang, The molecular origin and evolution of dim-light vision in mammals, Evolution 69 (11) (2015) 2995–3003.
[62] D.P. Anderson, D.S. Whitney, V. Hanson-Smith, A. Woznica, W. Campodonico-Burnett, B.F. Volkman, N. King, J.W. Thornton, K.E. Prehoda, Evolution of an ancient protein function involved in organized multicellularity in animals, eLife 5 (2016).
[63] P.K. Tan, J.E. Farrar, E.A. Gaucher, J.N. Miner, Coevolution of URAT1 and uricase during primate evolution: implications for serum urate homeostasis and gout, Mol. Biol. Evol. 33 (9) (2016) 2193–2200.
[64] A.K. Garcia, J.W. Schopf, S.-i. Yokobori, S. Akanuma, A. Yamagishi, Reconstructed ancestral enzymes suggest long-term cooling of Earth's photic zone since the Archean, Proc. Natl. Acad. Sci. U.S.A. 114 (18) (2017) 4619–4624.
[65] S. Akanuma, Y. Nakajima, S.i. Yokobori, M. Kimura, N. Nemoto, T. Mase, K.i. Miyazono, M. Tanokura, A. Yamagishi, Experimental evidence for the thermophilicity of ancestral life, Proc. Natl. Acad. Sci. U.S.A. 110 (27) (2013) 11067–11072.
[66] V.A. Risso, J.A. Gavira, D.F. Mejia-Carmona, E.A. Gaucher, J.M. Sanchez-Ruiz, Hyperstability and substrate promiscuity in laboratory resurrections of precambrian β-lactamases, J. Am. Chem. Soc. 135 (8) (2013) 2899–2902.
[67] V. Nguyen, C. Wilson, M. Hoemberger, J.B. Stiller, R.V. Agafonov, S. Kutter, J. English, D.L. Theobald, D. Kern, Evolutionary drivers of thermoadaptation in enzyme catalysis, Science 355 (6322) (2017) 289–294.
[68] J.A. Ugalde, Evolution of coral pigments recreated, Science 305 (5689) (2004) 1433-1433.
[69] D.M. Hillis, Taxonomic sampling, phylogenetic accuracy, and investigator bias, Syst. Biol. 47 (1) (1998) 3–8.
[70] D.J. Zwickl, D.M. Hillis, Increased taxon sampling greatly reduces phylogenetic error, Syst. Biol. 51 (4) (2002) 588–598.
[71] T.A. Heath, S.M. Hedtke, D. Hillis, Taxon sampling and accuracy of phylogenetic analyses, J. Syst. Evol. 46 (3) (2008).
[72] J. Bergsten, A review of long-branch attraction, Cladistics 21 (2) (2005) 163–193.
[73] B.R. Holland, D. Penny, M.D. Hendy, Outgroup misplacement and phylogenetic inaccuracy under a molecular clock—a simulation study, Syst. Biol. 52 (2) (2003) 229–238.
[74] A. Graybeal, Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47 (1) (1998) 9–17.
[75] S. Poe, Evaluation of the strategy of long-branch subdivision to improve the accuracy of phylogenetic methods, Syst. Biol. 52 (3) (2003) 423–428.
[76] D.D. Pollock, D.J. Zwickl, J.A. McGuire, D.M. Hillis, Increased taxon sampling is advantageous for phylogenetic inference, Syst. Biol. 51 (4) (2002) 664–671.
[77] K. Katoh, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, Nucleic Acids Res. 30 (14) (2002) 3059–3066.
[78] A. Loytynoja, N. Goldman, From the Cover: an algorithm for progressive multiple alignment of sequences with insertions, Proc. Natl. Acad. Sci. U.S.A. 102 (30) (2005) 10557–10562.
[79] M. Anisimova, G. Cannarozzi, D.A. Liberles, Finding the balance between the mathematical and biological optima in multiple sequence alignment, Trends Evol. Biol. 2 (1) (2010) 7.
[80] S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, Syst. Biol. 52 (5) (2003) 696–704.
[81] A.L. Bazinet, D.J. Zwickl, M.P. Cummings, A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0, Syst. Biol. 63 (5) (2014) 812–818.
[82] L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE, A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (1) (2015) 268–274.
[83] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (9) (2014) 1312–1313.
[84] F. Ronquist, J.P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models, Bioinformatics 19 (12) (2003) 1572–1574.
[85] N. Lartillot, H. Philippe, A bayesian mixture model for across-site heterogeneities in the amino-acid replacement process, Mol. Biol. Evol. 21 (6) (2004) 1095–1109.
[86] M.A. Suchard, B.D. Redelings, BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny, Bioinformatics 22 (16) (2006) 2047–2048.
[87] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, Evolution 39 (4) (1985) 783.
[88] B. Wróbel, Statistical measures of uncertainty for branches in phylogenetic trees inferred from molecular sequences by using model-based methods, J. Appl. Genet. 49 (1) (2008) 49–67.
[89] F. Abascal, R. Zardoya, D. Posada, ProtTest: selection of best-fit models of protein evolution, Bioinformatics 21 (9) (2005) 2104–2105.
[90] N.I. Bloch, J.M. Morrow, B.S.W. Chang, T.D. Price, SWS2 visual pigment evolution as a test of historically contingent patterns of plumage color evolution in warblers, Evolution 69 (2) (2015) 341–356.
[91] Z. Yang, PAML: a program package for phylogenetic analysis by maximum likelihood, Bioinformatics 13 (5) (1997) 555–556.
[92] B.G. Hall, Simple and accurate estimation of ancestral protein sequences, Proc. Natl. Acad. Sci. U.S.A. 103 (14) (2006) 5431–5436.
[93] K. Aadland, C. Pugh, B. Kolaczkowski, High-throughput reconstruction of ancestral protein sequence, structure, and molecular function, in: T. Sikosek (Ed.), Computational Methods in Protein Evolution, Humana Press, New York, 2018, pp. 135–170.
[94] J.M. Thomson, E.A. Gaucher, M.F. Burgan, D.W. De Kee, T. Li, J.P. Aris, S.A. Benner, Resurrecting ancestral alcohol dehydrogenases from yeast, Nat. Genet. 37 (6) (2005) 630–635.
[95] G.C. Finnigan, V. Hanson-Smith, T.H. Stevens, J.W. Thornton, Evolution of increased complexity in a molecular machine, Nature 481 (7381) (2012) 360–364.
[96] C. Bickelmann, J.M. Morrow, J. Du, R.K. Schott, I. van Hazel, S. Lim, J. Müller, B.S.W. Chang, The molecular origin and evolution of dim-light vision in mammals: Brief communication, Evolution 69 (11) (2015) 2995–3003.
[97] J.T. Bridgham, J. Keay, E.A. Ortlund, J.W. Thornton, Vestigialization of an allosteric switch: genetic and structural mechanisms for the evolution of constitutive activity in a steroid hormone receptor, PLoS Genet. 10 (1) (2014) e1004058.
[98] H. Bar-Rogovsky, A. Stern, O. Penn, I. Kobl, T. Pupko, D.S. Tawfik, Assessing the prediction fidelity of ancestral reconstruction by a library approach, Protein Eng. Des. Sel. 28 (11) (2015) 507–518.
[99] G.N. Eick, J.K. Colucci, M.J. Harms, E.A. Ortlund, J.W. Thornton, Evolution of minimal specificity and promiscuity in steroid hormone receptors, PLoS Genet. 8 (11) (2012) e1003072.
[100] M.J. DeNiro, S. Epstein, Mechanism of carbon isotope fractionation associated with lipid synthesis, Science 197 (4300) (1977) 261–263.
[101] J.P. Klinman, Kinetic isotope effects in enzymology, Adv. Enzymol. Relat. Area Mol. Biol. 46 (1978) 415–494.
[102] D.B. McNevin, M.R. Badger, S.M. Whitney, S. von Caemmerer, G.G.B. Tcherkez, G.D. Farquhar, Differences in carbon isotope discrimination of three variants of D-ribulose-1,5-bisphosphate carboxylase/oxygenase reflect differences in their catalytic mechanisms, J. Biol. Chem. 282 (49) (2007) 36068–36076.
[103] W.M. White, Geochemistry, John Wiley & Sons Inc., Hoboken, NJ, 2013.
[104] X. Zhang, D.M. Sigman, F.M.M. Morel, A.M.L. Kraepiel, Nitrogen isotope fractionation by alternative nitrogenases and past ocean anoxia, Proc. Natl. Acad. Sci. Unit. States Am. 111 (13) (2014) 4782–4787.
[105] S. Akanuma, S.-i. Yokobori, Y. Nakajima, M. Bessho, A. Yamagishi, Robustness of predictions of extremely thermally stable proteins in ancient organisms: Thermophilicity of ancient life, Evolution 69 (11) (2015) 2954–2962.
[106] L.P. Knauth, D.R. Lowe, High Archean climatic temperature inferred from oxygen isotope geochemistry of cherts in the 3.5 Ga Swaziland Supergroup, South Africa, Geol. Soc. Am. Bull. 115 (2003) 566–580.
[107] J.F. Kasting, M.T. Howard, K. Wallmann, J. Veizer, G. Shields, J. Jaffrés, Paleoclimates, ocean depth, and the oxygen isotopic composition of seawater, Earth Planet. Sci. Lett. 252 (1–2) (2006) 82–93.
[108] F. Robert, M. Chaussidon, A palaeotemperature curve for the Precambrian oceans based on silicon isotopes in cherts, Nature 443 (7114) (2006) 969–972.

[109] M.J. de Wit, H. Furnes, 3.5-Ga hydrothermal fields and diamictites in the Barberton Greenstone Belt–Paleoarchean crust in cold environments, Sci. Adv. 2 (2) (2016) e1500368-e1500368.
[110] A.L. Pey, D. Rodriguez-Larrea, S. Bomke, S. Dammers, R. Godoy-Ruiz, M.M. Garcia-Mira, J.M. Sanchez-Ruiz, Engineering proteins with tunable thermodynamic and kinetic stabilities, Proteins: Struct. Funct. Bioinf. 71 (1) (2008) 165–174.
[111] M. Goldsmith, D.S. Tawfik, Enzyme engineering by targeted libraries, Methods Enzymol. 523 (2013) 257–283.
[112] V.A. Risso, J.A. Gavira, E.A. Gaucher, J.M. Sanchez-Ruiz, Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins: consensus vs. Ancestral Proteins, Proteins: Struct. Funct. Bioinf. 82 (6) (2014) 887–896.
[113] M. Lehmann, C. Loch, A. Middendorf, D. Studer, S.F. Lassen, L. Pasamontes, A.P.G.M. van Loon, M. Wyss, The consensus concept for thermostability engineering of proteins: further proof of concept, Protein Eng. Des. Sel. 15 (5) (2002) 403–411.
[114] C. Kiss, J. Temirov, L. Chasteen, G.S. Waldo, A.R.M. Bradbury, Directed evolution of an extremely stable fluorescent protein, Protein Eng. Des. Sel. 22 (5) (2009) 313–323.
[115] D.L. Trudeau, M. Kaltenbach, D.S. Tawfik, On the potential origins of the high stability of reconstructed ancestral proteins, Mol. Biol. Evol. 33 (10) (2016) 2633–2641.
[116] S. Akanuma, Characterization of reconstructed ancestral proteins suggests a change in temperature of the ancient biosphere, Life 7 (3) (2017) 33.
[117] H.D. Holland, The oxygenation of the atmosphere and oceans, Philos. Trans. Roy. Soc. B 361 (2006) 903–915.
[118] H.J. Hofmann, Precambrian microflora, Belcher Islands, Canada; significance and systematics, J. Paleontol. 50 (6) (1976) 1040–1073.
[119] C. Klein, N.J. Beukes, J.W. Schopf, Filamentous microfossils in the early proterozoic transvaal supergroup: their morphology, significance, and paleoenvironmental setting, Precambrian Res. 36 (1) (1987) 81–94.
[120] B. Rasmussen, I.R. Fletcher, J.J. Brocks, M.R. Kilburn, Reassessing the first appearance of eukaryotes and cyanobacteria, Nature 455 (7216) (2008) 1101–1104.
[121] R.E. Blankenship, Early evolution of photosynthesis, Plant Physiol. 154 (2) (2010) 434–438.
[122] R. Couñago, S. Chen, Y. Shamoo, In vivo molecular evolution reveals biophysical origins of organismal fitness, Mol. Cell 22 (2006) 441–449.
[123] C.L. Worth, S. Gong, T.L. Blundell, Structural and functional constraints in the evolution of protein families, Nat. Rev. Mol. Cell Biol. 10 (2009) 709–720.
[124] T. Sikosek, H.S. Chan, Biophysics of protein evolution and evolutionary protein biophysics, J. R. Soc. Interface 11 (2014) 20140419.
[125] G. Fuchs, Alternative pathways of carbon dioxide fixation: insights into the early evolution of life, Annu. Rev. Microbiol. 65 (2011) 631–658.
[126] M.C. Weiss, F.L. Sousa, N. Mrnjavac, S. Neukirchen, M. Roettger, S. Nelson-Sathi, W.F. Martin, The physiology and habitat of the last universal common ancestor, Nat. Microbiol. 1 (2016) 16116.
[127] F.R. Tabita, Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective, Photosynth. Res. 60 (1) (1999) 1–28.
[128] R.M. McCourt, C.F. Delwiche, K.G. Karol, Charophyte algae and land plant origins, Trends Ecol. Evol. 19 (12) (2004) 661–666.
[129] A.D. Anbar, Y. Duan, T.W. Lyons, G.L. Arnold, B. Kendall, R.A. Creaser, A.J. Kaufman, G.W. Gordon, C. Scott, J. Garvin, R. Buick, A whiff of oxygen before the great oxidation event? Science 317 (5846) (2007) 1903–1906.
[130] J.W. Schopf, Disparate rates, differing fates: tempo and mode of evolution changed from the Precambrian to the Phanerozoic, Proc. Natl. Acad. Sci. U.S.A. 91 (15) (1994) 6735–6742.
[131] S.M. Whitney, R.L. Houtz, H. Alonso, Advancing our understanding and capacity to engineer nature's CO2-sequestering enzyme, Rubisco, Plant Physiol. 155 (2011) 27–35.
[132] M. Schidlowski, Carbon isotopes as biogeochemical recorders of life over 3.8 Ga of Earth history: evolution of a concept, Precambrian Res. 106 (2001) 117–134.
[133] M.F. Estep, F.R. Tabita, P.L. Parker, C. Van Baalen, Carbon isotope fractionation by ribulose-1,5-bisphosphate carboxylase from various organisms, Plant Physiol. 61 (1978) 680–687.
[134] K.M. Scott, M. Henn-Sax, T.L. Harmer, D.L. Longo, C.H. Frame, C.M. Cavanaugh, Kinetic isotope effect and biochemical characterization of form IA RuBisCO from the marine cyanobacterium Prochlorococcus marinus MIT9313, Limnol. Oceanogr. 48 (2003) 48–54.
[135] P.P. Wong, R.H. Burris, Nature of oxygen inhibition of nitrogenase from *Azotobacter vinelandii*, Proc. Nat. Acad. Sci. U.S.A. 69 (1972) 672–675.
[136] A.J. Milligan, I. Berman-Frank, Y. Gerchman, G.C. Dismukes, P.G. Faklowski, Light-dependent oxygen consumption in nitrogen-fixing cyanobacteria plays a key role in nitrogenase protection, J. Phycol. 43 (2007) 845–852.
[137] R.R. Eady, Structure – Function relationships of alternative nitrogenases, Chem. Rev. 96 (7) (1996) 3013–3030.
[138] J. Raymond, J.L. Siefert, C.R. Staples, R.E. Blankenship, The natural history of nitrogen fixation, Mol. Biol. Evol. 21 (3) (2004) 541–554.
[139] E.S. Boyd, A.D. Anbar, S. Miller, T.L. Hamilton, M. Lavin, J.W. Peters, A late methanogen origin for molybdenum-dependent nitrogenase, Geobiology 9 (3) (2011) 221–232.
[140] E.S. Boyd, T.L. Hamilton, J.W. Peters, An alternative path for the evolution of biological nitrogen fixation, Front. Microbiol. 2 (2011) 205.
[141] P. Rowell, W. James, W.L. Smith, L.L. Handley, C.M. Scrimgeour, ^{15}N discrimination in molybdenum- and vanadium-grown N_2-fixing *Anabaena variabilis* and *Azotobacter vinelandii*, Soil Biol. Biochem. 30 (14) (1998) 2177–2180.
[142] A.L. Zerkle, C.K. Junium, D.E. Canfield, C.H. House, Production of ^{15}N-depleted biomass during cyanobacterial N_2-fixation at high Fe concentrations, J. Geophys. Res. 113 (2008) G03014.
[143] J.B. Glass, F. Wolfe-Simon, J.J. Elser, A.D. Anbar, Molybdenum-nitrogen co-limitation in freshwater and coastal heterocystous cyanobacteria, Limnol. Oceanogr. 55 (2) (2010) 667–676.
[144] V. Beaumont, F. Robert, Nitrogen isotope ratios of kerogens in Precambrian cherts: a record of the evolution of atmosphere chemistry? Precambrian Res. 96 (1999) 63–82.
[145] E.E. Stüeken, R. Buick, B.M. Guy, M.C. Koehler, Isotopic evidence for biological nitrogen fixation by molybdenum-nitrogenase from 3.2 Gyr, Nature 520 (7549) (2015) 666–669.
[146] A.L. Zerkle, S.W. Poulton, R.J. Newton, C. Mettam, M.W. Claire, A. Bekker, C.K. Junium, Onset of the aerobic nitrogen cycle during the great oxidation event, Nature 542 (7642) (2017) 465–467.
[147] P.C. Dos Santos, Z. Fang, S.W. Mason, J.C. Setubal, R. Dixon, Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes, BMC Genomics 13 (2012) 162.
[148] A.K. Garcia, Ancient Biochemical Evolution of Nitrogenase through the History of Earth Oxygenation, 2018, American Geophysical Union Fall Meeting, Washingon D.C., USA, 2018.