

Metabolic Evolution of a Deep-Branching Hyperthermophilic Chemoautotrophic Bacterium

Rogier Braakman^{1*}‡, Eric Smith^{1,2}

1 Santa Fe Institute, Santa Fe, New Mexico, United States of America, **2** Krasnow Institute for Advanced Study, George Mason University, Fairfax, Virginia, United States of America

Abstract

Aquifex aeolicus is a deep-branching hyperthermophilic chemoautotrophic bacterium restricted to hydrothermal vents and hot springs. These characteristics make it an excellent model system for studying the early evolution of metabolism. Here we present the whole-genome metabolic network of this organism and examine in detail the driving forces that have shaped it. We make extensive use of phylometabolic analysis, a method we recently introduced that generates trees of metabolic phenotypes by integrating phylogenetic and metabolic constraints. We reconstruct the evolution of a range of metabolic sub-systems, including the reductive citric acid (rTCA) cycle, as well as the biosynthesis and functional roles of several amino acids and cofactors. We show that *A. aeolicus* uses the reconstructed ancestral pathways within many of these sub-systems, and highlight how the evolutionary interconnections between sub-systems facilitated several key innovations. Our analyses further highlight three general classes of driving forces in metabolic evolution. One is the duplication and divergence of genes for enzymes as these progress from lower to higher substrate specificity, improving the kinetics of certain sub-systems. A second is the kinetic optimization of established pathways through fusion of enzymes, or their organization into larger complexes. The third is the minimization of the ATP unit cost to synthesize biomass, improving thermodynamic efficiency. Quantifying the distribution of these classes of innovations across metabolic sub-systems and across the tree of life will allow us to assess how a tradeoff between maximizing growth rate and growth efficiency has shaped the long-term metabolic evolution of the biosphere.

Citation: Braakman R, Smith E (2014) Metabolic Evolution of a Deep-Branching Hyperthermophilic Chemoautotrophic Bacterium. PLoS ONE 9(2): e87950. doi:10.1371/journal.pone.0087950

Editor: Valerie de Crécy-Lagard, University of Florida, United States of America

Received: September 14, 2013; **Accepted:** January 5, 2014; **Published:** February 5, 2014

Copyright: © 2014 Braakman, Smith. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Parts of this work were performed under support from NSF FIBR grant nr. 0526747—The Emergence of Life: From Geochemistry to the Genetic Code. RB was further supported by an SFI Omidyar Fellowship at the Santa Fe Institute. ES thanks Insight Venture Partners for support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: One of the authors (Eric Smith) is supported by donations from Insight Venture Partners. This support places no restrictions on either topics or conclusions of any of the author's work. Insight Venture Partners has no interest, financial or otherwise, in any contents of this work. The support does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: rogier@santafe.edu

‡ Current address: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America.

Introduction

Metabolism lies at the heart of cellular physiology, acting as a chemical transformer between environmental inputs and components of biomass. Identifying rules and principles that underlie metabolic architecture can thus provide important insights into how basic properties of chemistry and physics constrain living systems. Of particular relevance to understanding the chemical history of the biosphere is the foundational layer of autotrophic metabolism, which fixes CO₂ and ultimately provides the ecological support to all forms of heterotrophy.

The merits of this view [1] were highlighted in a recent study on the early evolution of carbon-fixation pathways, which concluded that environmentally-driven innovations in this process underpin most of the deepest branches in the tree of life [2]. To extend our analysis of the early evolution of metabolism and of autotrophy, we present here a whole-genome reconstruction of the metabolic network of *Aquifex aeolicus*. *A. aeolicus* is a chemoautotroph, deriving both biomass and energy from inorganic chemical compounds, and is one of the deepest-branching and most thermophilic known bacteria [3]. Deep-branching clades restricted to hydrothermal

vents are generally considered to contain some of the most conservative metabolic features as a result of high degree of long-term stability provided by these environments [4].

While *A. aeolicus* has been the focus of substantial experimental efforts (see Ref. [5] for a review), it has not been characterized nearly as extensively as other model systems for which highly curated metabolic models exist. In addition, the inherent uncertainty of genome annotation from sequence alone [6,7], while overall significantly improving for next-generation methods [8], is compounded by the deep-branching position and extremophile lifestyle of this organism. Metabolic reconstruction protocols generally rely on heuristic rules to deal with the inevitable network “gaps” that result from misannotation or the presence of genes of unknown function. Such protocols tend to perform well in predicting basic aspects of phenotype, such as growth rate, particularly for well-studied organisms [9,10], but it is less clear what level of confidence to assign them when the focus is the evolution of specific metabolic sub-systems. Moreover, reconstructing an individual metabolic network requires

substantial effort and provides only a single “snapshot” of an evolutionary process that has played out over several billion years.

For these reasons we utilize phylometabolic analysis (PMA) [2] to guide the reconstruction of the metabolic network of *A. aeolicus* from its genome [11]. PMA generates trees of functional metabolic networks (i.e. phenotypes) by integrating metabolic and phylogenetic reconstructions. The power of PMA derives from a simple yet versatile constraint: the continuity of life in evolution. Since metabolic pathways are the supply lines of monomers from which all life is constructed, the continuity of life requires that at the ecosystem level *some* pathway to a given universal metabolite *must* always be complete in any evolutionary sequence across different parts of the tree of life. The distribution of metabolic genes in different pathways to given metabolites, within and across clades, thus informs the most likely completions in individuals, while distributions of pathways suggest the evolutionary sequences that connect them (see also Methods). We recently introduced PMA to reconstruct the evolutionary history of carbon-fixation, relating all extant pathways to a single ancestral form [2]. Here we show the versatility of this approach, using it to reconstruct the complete whole-genome metabolic network of an individual species, while further examining the evolutionary driving forces that have shaped the network.

As we will show, *A. aeolicus* synthesizes a significant fraction of its biomass through metabolic pathways that appear to represent conserved forms of the ancestral pathways to those metabolites. This is relevant in debates on the position of this organism within the tree of life. Initial phylogenetic studies based on 16S rRNA suggested that the Aquificales represent potentially the deepest branch within the bacterial domain [12,13]. Later studies of conserved insertion-deletions (indels) in a range of proteins led to the conclusion that Aquificales are instead a later branch more closely related to -proteobacteria [14], but this was subsequently found to be likely the result of substantial horizontal gene transfer (HGT) between these two clades [15]. It has since become clear that Aquificales and -proteobacteria represent the dominant clades of primary producers near hydrothermal vents [16], and ecological association is now understood to be a major driver of HGT [17]. Together this appears to have restored some consensus on the very deep-branching position of *A. aeolicus* [15,18], which is further supported by our analysis of its metabolism.

Innovations in Metabolic Evolution

Our analysis highlights three classes of innovations in metabolic evolution. One of these is gene duplication and divergence along a progression from low substrate-specificity to high substrate-specificity enzymes, driven by selection for improved kinetics. It has long been argued that early enzymes had broader substrate affinities than modern enzymes, with greater potential to promiscuously catalyze homologous reactions in earlier times [19–22]. Broad affinity of ancestral enzymes is thought to have facilitated evolutionary adaptation by providing a ‘background’ of biochemical pathways, initially proceeding at lower rates. If such background pathways produced advantageous products, they could then have been incorporated *en bloc* into metabolism [19], initially through increased enzyme expression levels and eventually through duplication and divergence leading to more specific enzymes [23].

While selection for improved kinetics has probably lowered the overall occurrence of promiscuous enzymes, metabolism maintains a substantial degree of enzyme promiscuity. For example, *E. coli* mutants in which an essential metabolic pathway was knocked out have been observed to recruit an alternate pathway from parts normally used for other functions to maintain growth [24]. In

addition to promiscuity in the binding of alternate substrates while local functional group transformation is preserved (substrate promiscuity), promiscuity can also occur through the catalysis of alternate reaction mechanisms (catalytic promiscuity). The form of promiscuity most frequently encountered in extant cells appears to be substrate promiscuity [25]. In general one might expect that the inherent “messiness” of enzymatic chemistry leads to a cost-benefit tradeoff in the evolution of substrate specificity, where complete specificity is difficult to achieve and moreover disadvantageous because it would decrease the capacity for future adaptation [23].

In particular one would expect this tradeoff to be different for core processes, where a higher mass flux can significantly amplify the benefits of improved kinetics, versus more peripheral processes that have lower mass flux. In keeping with this expectation, it is found that substrate promiscuity tends to increase toward the metabolic periphery [26], while reaction rate constants of enzymes tend to increase toward the metabolic core [27]. The idea that selection for kinetics has determined the degree of specificity of a pathway’s enzymes raises the intriguing possibility that prior to selection for increased substrate specificity, homologous reaction sequences could have initially been catalyzed by the same set of promiscuous enzymes [19,28,29], allowing earlier metabolisms with greater abundances of such homologous sequences to be controlled with smaller genomes.

We will discuss the evolution of several sub-networks in the metabolism of *A. aeolicus* that provide illustrations of these general principles. We will show that compared with later branching autotrophs *A. aeolicus* uses a greater abundance of repeated parallel chemical sequences catalyzed by enzymes with high sequence similarity, which could have initially been catalyzed by a smaller number of more promiscuous enzymes. We will further highlight how cost-benefit tradeoffs of improving kinetics have played out differently in core versus peripheral pathways.

A second class of innovations concerns the fusion of enzyme subunits into larger single-subunit enzymes. Enzyme fusion increases the effective concentrations of reactants at the active site for subsequent reactions within a sequence, thus increasing the throughput rate of the sequence [30]. In particular when intermediates within pathway sequences are not used elsewhere in metabolism, the fusion of associated enzymes would appear to have potentially little cost, but significant kinetic benefits for pathways with high mass-flux densities. Accordingly, studies have generally shown that gene fusion events significantly outnumber gene fission events in evolution [31,32]. Similarly, the organization of enzymes into multi-subunit enzymes, or possibly even super-complexes, can, in addition to facilitating novel reaction mechanisms, also be considered to improve the kinetics of metabolism by increasing effective concentrations of intermediates. While this is not a central focus of the current study, we will highlight several key reaction sequences that in *A. aeolicus* are catalyzed by multi-subunit enzymes known to be fused in later branching organisms, reflecting *Aquifex*’s more primitive metabolism.

A final important class of innovations occurs at the level of pathways. In some cases organisms may have access to multiple different pathways which produce certain metabolites, with different pathways’ having different unit costs of required ATP hydrolysis. Recent work has suggested that lowering the ATP cost, thereby increasing overall thermodynamic efficiency, of pathways involved in CO₂-fixation was a major evolutionary driving force that resulted in several major early branches in the tree of life [2]. Here we show additional divergences that appear to be driven by increasing thermodynamic efficiency.

These three classes of innovations are all different expressions, at different levels, of a more general evolutionary tradeoff, in

which either the efficiency or the rate of growth is maximized. For example, for heterotrophic organisms it is observed that in the presence of resource competition cells will use fermentative metabolic modes and maximize the rate of ATP production, while in the absence of competition cells will use respiration to maximize the yield of ATP production [33]. Because ATP hydrolysis drives biosynthesis, increasing its production rate will increase growth rate, while increasing ATP yield will increase growth efficiency.

A second aspect of metabolism where this growth rate/efficiency tradeoff is expressed, is not in the production of ATP, but in the structure of the biosynthetic machinery driven by its subsequent hydrolysis. For example, improving kinetics of metabolic sequences through increased substrate specificity or gene fusion contributes to improving growth rate, while lowering ATP cost by choosing alternative pathways contributes to improving efficiency. The relative mass contributions of different pathways or reactions are likely a critical determinant in the balance of benefit vs. cost of such innovations in different parts of metabolism.

Cost-benefit tradeoffs in the structure of biosynthesis are likely to be of particular importance in autotrophic metabolism. Because (some) heterotrophs can switch between fermentative and respiratory strategies for given organic inputs, they have access to much larger variability in the rate of ATP production than do autotrophs. Autotrophs generally use purely respiratory metabolic modes to interconvert pairs of inorganic compounds and are thus more constrained on the ATP production side. Since autotrophs form the metabolic basis for ecosystems, the same will tend to hold for aggregate metabolic networks at that level. The biosphere is globally autotrophic, and especially on the longest time-scales of selection we may thus expect that just as new inorganic energy sources were being explored to increase total inputs to ecosystems, both the rate and the efficiency of metabolic processes were simultaneously being optimized where possible. Even if individuals have access to, and can switch between alternate strategies and lifestyles, both those individuals as well as the ecosystem to which they belong would benefit from improved kinetics and efficiency of their metabolisms. The present reconstruction of the metabolism of *A. aeolicus* provides an excellent testbed to start cataloging and exploring these ideas.

Methods

Principles of Phylometabolic Analysis

The basic principles of phylometabolic analysis (PMA) [2] are outlined in Fig. 1. The effectiveness and flexibility of PMA for studying metabolic evolution arises from the strongly synergistic potential of integrating metabolic and phylogenetic constraints into a single approach. For example, if an individual annotated genome leads to a putative metabolic network that is not viable due to gaps in all known pathways that connect given sub-networks, then without experimental evidence it may be very unclear how to proceed in the curation process. As shown in panel A, placing that same putative network in phylogenetic context and comparing metabolic gene profiles within and across clades, often clearly suggests the proper gap-fills. We will show many such examples of the ways PMA guides the reconstruction process for the metabolic network of *A. aeolicus*.

The same genomic surveys that help interpret the character of individual phenotypes also produce distributions of metabolic pathways across the tree of life, as shown in panel B. PMA exploits universal topological bottlenecks within metabolism – metabolites through which all pathway alternatives in a given sub-network

must pass to allow biosynthesis. Imposing the evolutionary constraint that all extant and ancestral metabolic states supply those bottleneck metabolites allows us to transform the distribution of pathways into specific sequences of pathway innovations. These sequences can be represented as phylometabolic trees, as shown on the right in panel B. As can be seen, the constraint of continuous metabolite production requires evolutionary intermediates in which two pathways are co-present before one may replace the other.

Phylometabolic analysis imposes a network-level constraint on constellations of genes, which is not imposed in standard phylogenies using only gene presence/absence or similarity data. However, it is important to recognize that constraints from metabolic stoichiometry alone cannot distinguish whether the genes that make up a network are all contained within the genome of individual organisms, or are distributed within syntrophic communities. Although the distributions of genes that we use as a basis for evolutionary reconstructions come from comparing individual genomes, this does not imply that the gene constellations for reconstructed ancestral states of a given metabolic sub-network were likewise contained within a single species rather than within communities.

The nodes and links in a phylometabolic tree reflect the serial and parallel dependencies among metabolic innovations at the network level. Whether the serial dependencies, represented by internal nodes, are due to branching among individual lineages, or due to branching among tightly coupled consortia of cells, is not determined by an analysis using only metabolic genes. Such distinctions require more comprehensive analyses involving transporters and perhaps regulatory genes for organisms spanning the tree. We therefore generally withhold interpretations about individuality for internal nodes when reconstructing a phylometabolic tree. Nevertheless, the genomes of individual species may still contain most or all of the network contained within internal nodes of a phylometabolic tree. This is particularly true for autotrophs (such as *A. aeolicus*), which generate all biomass from inorganics and are in a sense “ecosystems within individuals”, and are thus useful model systems for reconstructing the long-term evolution of metabolism.

Finally, because nodes are now metabolic phenotypes, comparing various aspects of pathways, including the reaction types they contain, their ATP costs, and the structure of their regulation, can identify clear physical-chemical driving forces underlying divergences [2]. For *A. aeolicus* this helps us identify the key evolutionary driving forces that have shaped its metabolism.

Reconstructing the Metabolism of *A. aeolicus*

Our reconstruction of a whole-metabolism model for *A. aeolicus* is based on an initial network obtained from the Model SEED server [34], an automated pipeline that generates genome-scale metabolic models directly from genome annotations. After first modifying the nutrient and biomass compositions of the model to accurately capture the boundary conditions that define the overall *A. aeolicus* phenotype (see Text S1 for additional details), the internal network was curated, using criteria of phylometabolic consistency to evaluate proposed completions of key network gaps.

In practice, PMA was implemented by surveying the annotated genomes of a large number of archaea and bacteria across the tree of life at the online Uniprot database [35]. Other repositories may be used, but we chose Uniprot because it derives from the manually curated Swissprot database, which is known to have among the lowest error rates in gene annotation [7]. We then performed an exhaustive and redundant search for all relevant genes, using naming variants and EC classification numbers,

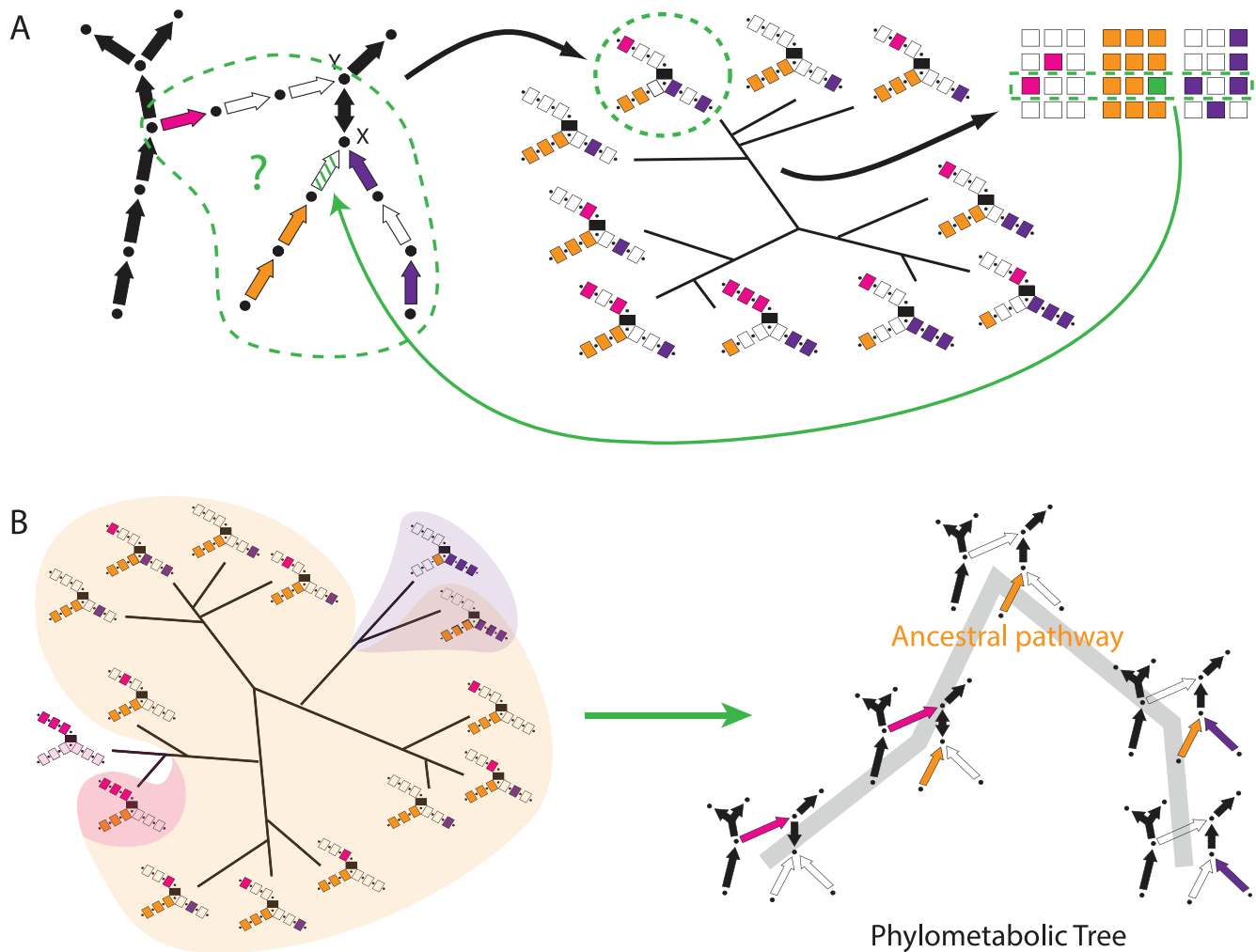


Figure 1. Principles of Phylometabolic Analysis. Panel A shows how phylogenetic distributions of pathways helps interpret the curation of individual metabolic networks. In this case comparison of metabolic gene profiles suggests the orange pathway represents the correct completion. Panel B in turn shows how pathway distributions in turn also suggest evolutionary sequences. Imposing continuity of metabolite production at the ecosystem level allows us to represent those sequences as phylometabolic trees, in which each node represents a functional phenotype with an explicit internal chemical structure.
doi:10.1371/journal.pone.0087950.g001

which code for the enzymes across a set of metabolic pathways that define a metabolic sub-system. When all enzymes of a metabolic pathway are identified within a single genome, then that pathway is counted as present in that species, and totals are tabulated across clades. The high rate at which new genomes are being sequenced is reflected in the fact that the number of members in clades differs in our tabulations for different sub-systems, even though analyses were in some cases performed only a few weeks apart.

To further correct for misannotated or unannotated genes, the above searches were complemented by additional BLAST searches using the built-in capabilities on the Uniprot website. As we will highlight with several examples, in many cases BLAST searches will identify groups of genes coding for the same enzyme at the clade level. Especially when organisms contain closely related enzymes that because of high sequence similarity are misannotated (often as the same gene), comparisons using both sequence similarity and enzyme length will often separate those enzymes into clear groups at the clade level. This in turn increases the chances that individual enzymes within those clade-level

groups have been experimentally studied within an individual member of that clade, helping to anchor the functional annotation of each of the enzyme groups.

Results and Discussion

Aquifex aeolicus provides an interesting reference point for minimal deep-branching hyperthermophilic bacterial metabolism. Our reconstruction for this organism contains 756 reactions and 729 metabolites (see File S1 for the whole-genome metabolic network in SBML format [36]). In comparison, the recently reconstructed network of *Thermotoga maritima* – like *A. aeolicus* thought to be one of the deepest-branching and most thermophilic bacteria [18] – contains 562 reactions and 503 metabolites [37]. Most of the difference in size results from the ways fatty acid metabolism is represented. In the *T. maritima* network, lipid metabolism is represented using many composite reactions, resulting in only 27 reactions and 27 metabolites in this sub-network [37]. By contrast, we explicitly represent each reaction in lipid synthesis, and because the lipid composition of *A. aeolicus* is diverse, this results in nearly 200 reactions and 200 metabolites

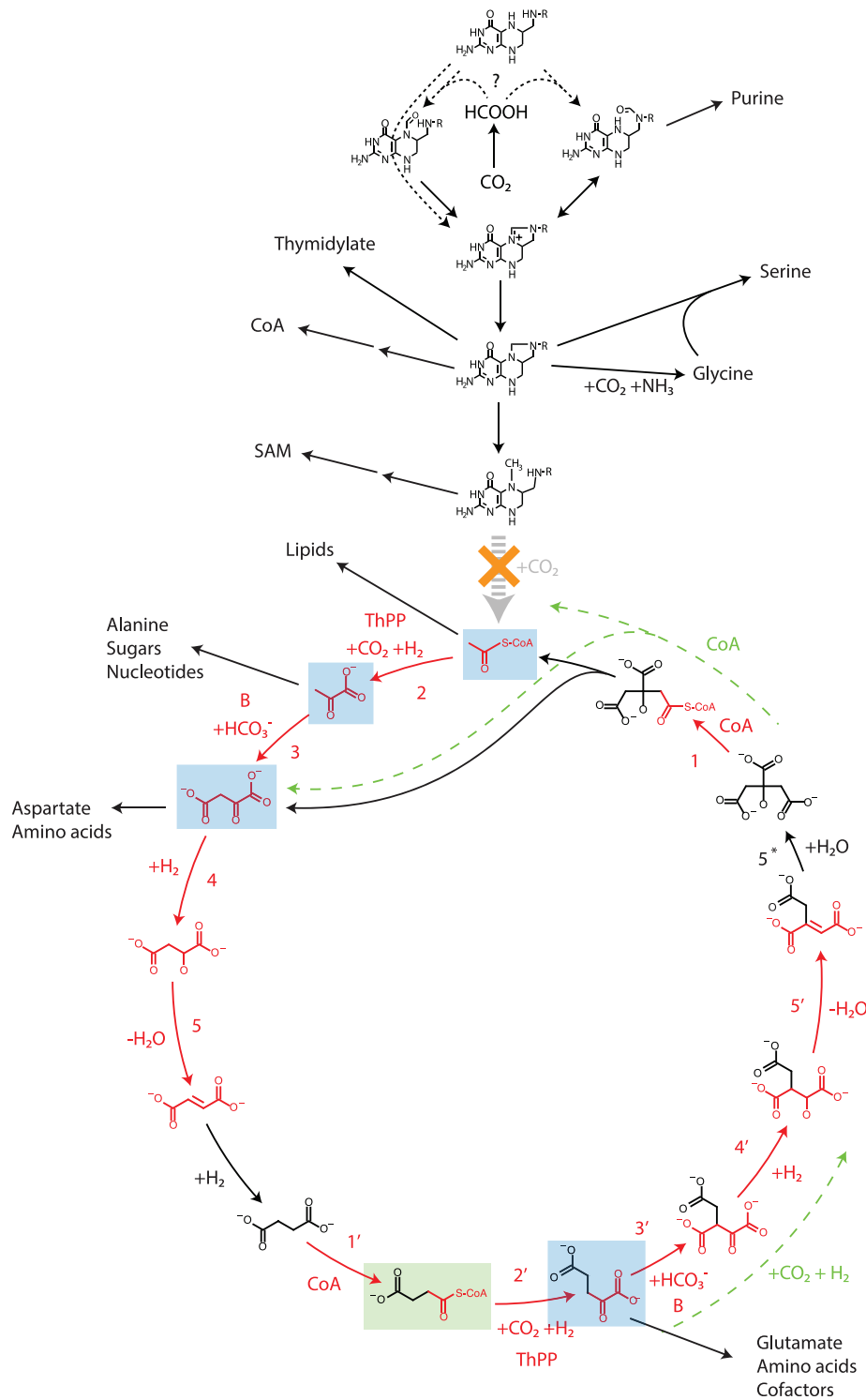


Figure 2. Carbon-fixation in *Aquifex aeolicus*. The main fixation pathway is the reductive citric acid (rTCA) cycle, from which most anabolic pathways are initiated. Reductive folate chemistry is a secondary fixation pathway from which an additional small set of anabolic pathways is initiated (R = aminobenzoate-derived side chain). Whether formate attachment occurs at the N¹⁰ or N⁵ position of THF remains to be elucidated (see text). Relative to the reconstructed root of carbon-fixation, in which the rTCA cycle and Wood-Ljungdahl pathway are fully integrated [2], this hybrid strategy employed by *A. aeolicus* lacks only the grey-dashed reaction (acetyl-CoA synthesis). Molecules highlighted in blue represent the “pillars of anabolism”, TCA intermediates from which the vast majority of anabolic pathways have been initiated throughout evolution [42]. Highlighted in green is succinyl-CoA, which forms a precursor to pyrroles through a later derived pathway in some organisms (but not *A. aeolicus*). Highlighted in red are reaction sequences involving the same local functional group transformation that in *A. aeolicus* are catalyzed by closely related enzymes in both halves of the rTCA cycle. Green dashed arrows highlight alternate pathway sequences catalyzed by a single enzyme in other clades.

doi:10.1371/journal.pone.0087950.g002

(see also the section on lipid biosynthesis below). However, fatty acid biosynthetic machinery is highly flexible and modular, with chains of diverse length and substitution patterns generated by a very small set of proteins [38]. It should further be noted that *T. maritima* in fact also has a diverse mixture of chain length and substitution patterns in its lipid content [39], but that the simpler composition of *E. coli* was chosen in defining its biomass for reconstruction [37]. Thus, taking this representation difference into account, the metabolic networks of *A. aeolicus* and *T. maritima* may be reasonably representative of minimal bacterial metabolism found near hydrothermal vents. Here we examine in detail the metabolic evolution of *A. aeolicus*, paving the way for future studies that may help disentangle the evolutionary relatedness between Aquificales and Thermotogales.

Our primary goal was to accurately reconstruct the pathway structure of the metabolism for the purpose of understanding its evolution. We therefore did not significantly modify the biomass vector of the model (with fatty acids as the major exception), or general assignments of genes to enzymes from the initial model obtained from the SEED server. This work should thus be considered a qualitative first step toward building a comprehensive computational organism model for *A. aeolicus*. As additional experimental data become available, including in particular detailed data on the biomass composition under different growth conditions, they will provide further benchmarks to make this model increasingly quantitative. Nevertheless, this reconstruction contains a wealth of data on basic principles of metabolic evolution, as we will discuss in detail, starting from core carbon-fixation.

Carbon-fixation and the Initiation of Anabolism

Aquifex aeolicus uses a unique and previously unrecognized form of carbon-fixation. It has been known for more than two decades that *A. aeolicus* uses the reductive citric acid (rTCA) cycle to fix most of its carbon [40,41]. However, our study on the evolution of carbon-fixation led us to conclude that in parallel this organism uses an incomplete form of the reductive acetyl-CoA (Wood-Ljungdahl, WL) pathway to produce a small sub-set of biomass components [2]. This hybrid form of carbon-fixation was recognized through a broad survey of the biosynthetic pathways leading to glycine and serine. We identified substantial gaps in each of the conventionally recognized pathways to these compounds across many deep-branching clades. Instead these organisms often possess the complete gene complement for an alternate folate based reductive C₁ pathway that is also used as part of WL. Indeed, the suggested existence of widespread hybrid carbon-fixation strategies using a partial WL pathway was one of the key insights that allowed the reconstruction of a phylometabolic tree of carbon-fixation through fully autotrophic intermediates [1,2].

The complete carbon-fixation strategy reconstructed for *A. aeolicus* is shown in Fig. 2. The major fraction of carbon is fixed through the rTCA cycle, producing its intermediates acetyl-CoA, pyruvate, oxaloacetate, and α -ketoglutarate (highlighted in blue), from which anabolic pathways to most components of biomass subsequently radiate [42]. In parallel, reductive folate chemistry transforms CO₂ into C₁ units of different oxidation states, from which a small number of additional anabolic pathways are initiated. Formyl groups are used in the production of purine, methylene groups are used in the production of glycine and serine, as well as thymidylate and coenzyme A, while methyl groups are donated to S-adenosyl methionine (SAM), which mediates methyl-group chemistry [43]. This combined carbon-fixation strategy thus lacks only a single reaction relative to the reconstructed root of

carbon-fixation, in which the rTCA cycle and the WL pathway are fully integrated [2]. The *A. aeolicus* metabolic phenotype is lacking only the final synthesis of acetyl-CoA within WL, which is catalyzed by one of the most oxygen-sensitive enzymes in the biosphere [44].

The specific form of formate uptake by folate remains to be elucidated. *A. aeolicus* lacks the gene for 10-formyl-THF (tetrahydrofolate) synthetase, which catalyzes the attachment of formate at the N¹⁰ position of THF in the acetogenic version of WL [43]. However, a broad genomic survey across deep branches of the tree of life nonetheless strongly supports the use of a partial WL pathway over other alternatives to produce glycine and serine in *A. aeolicus* [2].

Formate incorporation in this organism is most likely catalyzed by an unrecognized N¹⁰-uptake protein, or through an unrecognized sequence involving attachment of formate at the N⁵ position of THF [2]. This alternate route was suggested by the very wide distribution across the tree of life of a gene for an ATP-dependent 5-formyl-THF cycloligase, present in the genomes of many organisms that like *A. aeolicus* appear to use a partial WL pathway to produce glycine and serine, but which lack the (also ATP-dependent) 10-formyl-THF synthetase. 5-formyl-THF cycloligase catalyzes the cyclization of 5-formyl-THF to methenyl-THF and has so far only been suggested to be part of a futile cycle, but its precise physiological role has long remained unclear [45,46]. The suggested use of an N⁵ uptake route for formate is appealing not only for these reasons, but also because it would provide an evolutionary intermediate between N⁵ formate uptake in the methanogenic version of WL and N¹⁰ formate uptake in the acetogenic version of WL.

While it has been recognized that most anabolic pathways originate from TCA intermediates, there exist some variation in this set of intermediates across clades. In particular, many organisms derive pyrroles (precursors to the cofactor family that includes heme and chlorophyll) from α -ketoglutarate through what is known as the C5 pathway, while α -proteobacteria and mitochondria derive pyrroles from succinyl-CoA [47]. This had previously already led to the suggestion that pyrrole synthesis from succinyl-CoA is a later evolutionary innovation [48], which is confirmed by genomic surveys. Table 1 shows that the pathway from succinyl-CoA is almost completely absent from deep-branching bacterial and archaeal clades, while the C5 pathway is nearly universally distributed. Thus, across most of the tree of life, and for most of metabolic evolution, carbon-fixation has fed into anabolism through only 4 TCA intermediates.

Evolution of the rTCA cycle. Variants in the form of the rTCA cycle used within the Aquificales provide insights into the evolutionary driving forces that have shaped this pathway. Recent studies in *Hydrogenobacter thermophilus*, which together with *A. aeolicus*

Table 1. Distribution of entry sequences to pyrrole biosynthesis.

	C5	Aminolevulinate
Domain		
Archaea (133)	108	0
Bacteria* (345)	290	5

Notes:* Includes Aquificales, Thermotogales, Firmicutes (Bacillales+Clostridia), Chloroflexi, Chlorobiales, Cyanobacteria, Nitrospirae, Deinococcales, Verrucomicrobia, Planctomycetes.

doi:10.1371/journal.pone.0087950.t001

belongs to the Aquificaceae family within Aquificales, identified novel enzymes for several rTCA reactions. The cleavage of citrate to acetyl-CoA and oxaloacetate, and the reductive carboxylation of α -ketoglutarate to isocitrate, each conventionally recognized as single enzyme reactions, are in *H. thermophilus* both catalyzed in two steps by distinct enzymes also found in the *A. aeolicus* genome (see Fig. 2) [49–52]. This Aquificaceae variant of the rTCA cycle has an increased degree of symmetry, and reveals previously unrecognized homology relations in enzymes, which recapitulate the similarity in the local-group chemistry of their substrates.

The newly discovered citryl-CoA synthase and α -ketoglutarate biotin carboxylase enzymes catalyze local functional group chemistry that is homologous to their counterparts succinyl-CoA synthase and pyruvate biotin carboxylase (reactions 1, 1' and 3, 3' in Fig. 2), respectively. Moreover, both sets of homologous enzymes have high sequence similarity, and were originally annotated as the same enzymes [49,51]. Similarly, pyruvate and α -ketoglutarate ferredoxin:oxidoreductase also catalyze homologous chemistry (reactions 2, 2'), and due to high sequence similarity were again annotated as the same enzyme in *A. aeolicus* [11].

These observations are striking in light of the discussion on metabolic innovations in the introduction. A hypothesis of a minimal ancestral metabolism with promiscuous enzymes catalyzing homologous reactions in parallel, is consistent with a highly symmetric rTCA variant used by members of the Aquificaceae. Aquificales are the deepest-branching clade using this pathway to fix carbon [53], and as mentioned their exclusive association with hydrothermal vents may result in some of the most conservative metabolic features [4].

Indeed, the new enzymes identified in the Aquificaceae have been argued to represent the ancestral rTCA enzymes [41,50,54]. The single step ATP citrate lyase found in most rTCA bacteria is suggested to have arisen through a gene fusion of the second sub-unit of citryl-CoA synthase with citryl-CoA lyase [41,49,50], while the combination of α -ketoglutarate biotin carboxylase plus isocitrate dehydrogenase (ICDH) is suggested to have been replaced by a single ICDH with increased substrate specificity [54].

To further explore the evolutionary driving forces that underly the evolution of the rTCA cycle, we used *H. thermophilus* enzymes as a benchmark to survey the genomes of all Aquificales, as well as other clades using the rTCA cycle. Table 2 shows the distributions of each of the newly identified high homology rTCA enzymes across all Aquificale genomes.

Two clear groups of Aquificales are distinguished by the rTCA variants they use. All members of the Aquificaceae possess two biotin carboxylase (BC) and two CoA synthase (CS) enzymes. By contrast, all member of the Hydrogenothermaceae and Desulfurobacteriaceae possess only single copies of these enzyme types, except *S. azorensis*, which possess two BC enzymes. In addition, in terms of both sequence similarity and length, both sub-units of the single BC enzyme in Hydrogenothermaceae and Desulfurobacteriaceae best match pyruvate (rather than α -ketoglutarate) BC in Aquificaceae. Similarly, the single CS enzyme in these families best matches succinyl-CoA synthase in Aquificaceae. Finally, all Hydrogenothermaceae and Desulfurobacteriaceae are known to possess ATP citrate lyase, while all except *S. azorensis* possess an ICDH enzyme that is significantly larger than the version found in Aquificaceae. This difference in length is consistent with the suggestion that an increased substrate specificity of the ICDH allowed it to supplant the combined function of the BC and more primitive ICDH enzymes [54]. This survey shows how sequence length of enzymes can be a useful secondary source of information,

after sequence similarity, in functional annotation. Thus, we conclude that all Aquificaceae possess the highly symmetric rTCA variant shown in Fig. 2, while other Aquificale families mostly possess the more conventional rTCA variant.

Clear evolutionary driving forces can be identified that connect the different extant forms of the rTCA cycle. The gene fusion within ATP citrate lyase increases the effective concentration of citryl-CoA in the subsequent cleavage reaction, and can thus be understood to improve the kinetics of the rTCA cycle. This gene fusion would appear to have low cost, not requiring evolution of an additional compensatory pathway to release citryl-CoA as a free intermediate, since it is not used elsewhere in metabolism.

Detailed thermodynamic analyses have in turn identified two classes of rTCA reactions that in isolation would require ATP hydrolysis to proceed: carboxylation reactions and carboxyl reduction reactions [55]. These costs can be avoided in a number of ways. Coupling carboxyl reductions to subsequent carboxylations through thioester intermediates (e.g. succinyl- or acetyl-CoA), allows the combined sequence to be driven by a single ATP hydrolysis, while coupling endergonic to subsequent exergonic reactions can eliminate the ATP cost of the combined sequence [55]. The replacement of the combination of α -ketoglutarate BC and the lower-specificity ICDH with a single high-specificity ICDH falls into this latter category. The α -ketoglutarate BC reaction in Aquificaceae costs 1 ATP per carbon incorporated [51], while the subsequent reduction to isocitrate is highly exergonic, resulting in a nearly-reversible sequence without ATP cost when the reactions are coupled [42,55,56]. This adaptation can thus be seen as driven by improving the thermodynamic efficiency of the rTCA cycle.

An additional factor that could have further increased the ATP savings of using a single higher-specificity ICDH to generate isocitrate from α -ketoglutarate is that like other β -ketoacids oxalosuccinate is subject to spontaneous decarboxylation. The lifetime of oxaloacetate (also a β -ketoacid) in water may be estimated at about 3 min at 90°C (Figure based on $\Delta H^\ddagger=17.2\text{kcal/mol}$ and $k_{25^\circ\text{C}}\approx 2.8\times 10^{-5}\text{s}^{-1}$ from Ref. [57].), and oxalosuccinate is more unstable. Metal ions and amines are known, at least for oxaloacetate, to further enhance the rate of spontaneous decarboxylation [57]. The decarboxylation of oxalosuccinate back to its precursor α -ketoglutarate would introduce a futile cycle that may have raised the average cost of the forward carboxylation reaction above 1 ATP.

Surveys in the genomes of other clades using the rTCA cycle (data not shown) are consistent with these suggested adaptations. Chlorobiales, a later branching photoautotrophic clade in which the rTCA cycle was originally described [58], “*Candidatus Nitrospira defluvii*”, a member of the Nitrospirae in which the rTCA cycle was recently discovered through metagenome analysis [59], as well as *Sulfurimonas*, an ϵ -proteobacterial family that uses the rTCA cycle [60], all use the less-symmetric rTCA variant. Their genomic features associated with the rTCA cycle are similar to those of the Hydrogenothermaceae and Desulfurobacteriaceae. All of these species possess a gene for ATP citrate lyase, have single copies of BC and CS enzymes, and have ICDH enzymes with a length of around 730–740 amino acid residues.

Together these observations can be used to reconstruct a phylometabolic tree of rTCA variants, which in turn represents a branch of increased resolution on a more general phylometabolic tree of carbon-fixation [2]. This tree is shown in Fig. 3. The hypothesized root node, shown on the left, consists of the symmetric rTCA variant, but catalyzed by a single set of of enzymes for both arcs. Duplication and divergence toward greater specificity enzymes through selection for improved kinetics then

Table 2. Distribution of rTCA enzymes in Aquificales.

Family - Species	ICDH	BC - LS		BC - SS		CS - LS		CS - SS	
		PYR	AKG	PYR	AKG	SUC-	CIT-	SUC-	CIT-
I <i>T. thermophilus</i>	421	617 (100)	652 (100)	475 (100)	472 (100)	383 (100)	429 (100)	293 (100)	344 (100)
<i>A. aeolicus</i>	426	614 (81)	655 (78)	477 (76)	472 (82)	385 (77)	436 (78)	305 (84)	368 (84)
<i>T. Albus</i>	419	614 (81)	653 (90)	475 (83)	472 (91)	382 (81)	432 (87)	295 (87)	341 (89)
<i>H. sp Y04AAS1</i>	421	619 (69)	638 (74)	475 (69)	472 (79)	382 (74)	423 (78)	291 (80)	329 (77)
II <i>S. azorensis</i>	421	614 (65)	647 (74)	482 (67)	472 (73)	388 (63)	** (31)	293 (72)	** (39)
<i>P. marinus</i>	747	613 (67)	** (50)	478 (64)	** (56)	388 (65)	** (31)	292 (69)	** (38)
<i>S. sp Y03AOP1</i>	746	616 (66)	** (47)	476 (66)	** (56)	389 (64)	** (30)	293 (72)	** (39)
III <i>D. thermolithotrophum</i>	735	616 (57)	** (50)	472 (58)	** (54)	389 (62)	** (31)	300 (70)	** (37)
<i>T. ammonificans</i>	735	618 (59)	** (49)	472 (59)	** (54)	389 (63)	** (30)	300 (70)	** (37)

Numbers in each column represent the length of the enzyme in terms of amino acid residues, and the numbers in parentheses are the sequence similarities relative to the *T. thermophilus* version of those enzymes. Entries ** are for species that have only one copy of that enzyme class, which are aligned with the type to which it has greatest sequence similarity (sequence similarity for alternate alignment also shown for comparison). Families: I - Aquificaceae, II - Hydrogenothermaceae, III - Desulfurothermaceae. Abbreviations: ICDH, isocitrate dehydrogenase; BC, biotin carboxylase; -CoA synthase; LS, large subunit; SS, small subunit; PYR, pyruvate; AKG, α -ketoglutarate; SUC-, succinyl-; CIT-, citryl.

doi:10.1371/journal.pone.0087950.t002

gives rise to the symmetric rTCA variant found in Aquificaceae, while gene fusion in ATP citrate lyase combined with an increased substrate specificity of ICDH gives rise to a second divergence that leads to the conventional rTCA cycle. A comparison of ATP citrate lyase genes suggested that Aquificale families using it obtained this enzyme through HGT from ϵ -proteobacteria, and that the Chlorobiales version represents the ancestral ATP citrate lyase [41]. Thus, the initial divergence between rTCA variants may have been between Aquificales and Chlorobiales, with two of

the Aquificale families later joining the conventional rTCA group through HGT, as a result of ecological association with ϵ -proteobacteria.

While not all five homologous enzymes in the symmetric Aquificaceae rTCA variant show the same sequence similarity as the first three reactions (1–3), we argue this is consistent with the suggested evolutionary scenario. Reactions 1–3 represent thermodynamic bottleneck reaction that in isolation would require ATP hydrolysis [55], and are catalyzed by elaborate multi-subunit

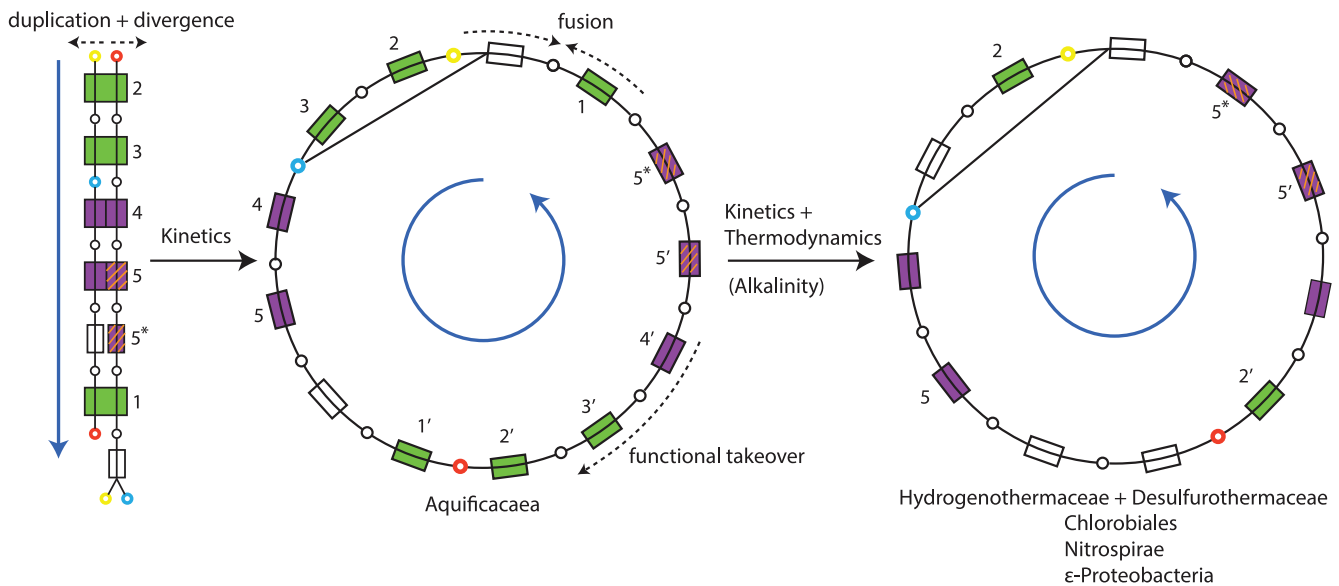


Figure 3. Phylometabolic tree showing the evolution of the rTCA cycle. A combination of improving kinetics (which increases growth rate) and improving thermodynamics (which increases growth efficiency) explains both divergences. For the first divergence, duplication and divergence toward higher substrate specificity of enzymes improves kinetics. For the second divergence, replacing enzymes 3' + 4' by a higher specificity version of enzyme 4 removes a cost of 1 ATP hydrolysis in fixing CO₂, while fusion of one sub-unit of enzyme 1 with enzyme for the subsequent cleavage reaction improves kinetics. Green boxes represent homologous reactions catalyzed by enzymes with high sequence similarity, while purple boxes represent homologous reactions catalyzed by members of the same enzyme families. Reactions 5' and 5* are catalyzed by the same enzyme. Differences in sequence divergence between green and purple enzymes may reflect differences in complexity of the reactions, see text for further discussion. The yellow node represents acetyl-CoA, the blue node represents oxaloacetate, and the red node represents succinyl-CoA. Dark blue arrows indicate the direction of mass through pathways.

doi:10.1371/journal.pone.0087950.g003

enzymes that contain complex metal-centers and/or complex organic cofactors, and belong to small and highly conserved enzyme families. By contrast, the hydrogenation and dehydration reactions (4, 5) that follow these complex reactions are catalyzed by simpler enzymes belonging to highly diversified enzyme families used throughout metabolism. The selection pressure to conserve enzyme sequence similarity is therefore likely to have been lower, and either (or both) of these enzymes could also have been more easily replaced by other members of this family. This breakdown between “easy” and “hard” chemistries has been recognized as a general constraint in the early evolution of carbon-fixation pathways, with innovations primarily occurring through the emergence of new sequences of easier chemistries that connect harder bottleneck reactions [1].

Having both arcs catalyzed by single sets of promiscuous enzymes in the root node would likely have resulted in slower growth for organisms using such a strategy, but it would also have considerably lowered the overhead for early regulatory machinery. If both rTCA arcs could have been catalyzed by the same set of single enzymes, the whole cycle would require only 7 total enzymes for catalysis: the 5 homologous reactions plus the reduction of malate to fumarate, and the cleavage of citryl-CoA, both of which do not have analogs in the opposite arc. It has previously been argued that a fully connected rTCA+WL network was selected as the root of carbon-fixation, because its topology would have provided the most robust form of network autocatalysis for earlier eras of life [1,2]. The present analysis suggests that an additional reason the rTCA cycle could have been privileged as a carbon fixation pathway capable of initiating the first cellular life, over alternate autocatalytic carbon-fixation pathways observed today [53,61], is that its greater symmetry permitted it to emerge with simpler catalytic support and regulatory structure.

Each of the divergences in the tree of rTCA variants can be understood in terms of a cost-benefit tradeoff in metabolic innovations. Duplication and divergence toward greater substrate specificity of the enzymes catalyzing the two arcs would have improved the kinetics of the pathway, and thus the growth rate of organisms using it. While the genome expansion due to these duplications would have had increased cost associated to it, this was apparently more than offset by the benefit of increasing the kinetics of a pathway through which most cellular carbon is incorporated. We will discuss a different case in the biosynthesis of branched chain amino acids in the next section, where a lower mass flux appears to have shifted this balance away from favoring duplication and divergence.

For the second divergence, the fusion of the citryl-CoA synthase and citryl-CoA lyase would have improved the kinetics of the rTCA cycle, while the elimination of the ATP dependent α -ketoglutarate BC would have improved the thermodynamic efficiency of the rTCA cycle. A secondary effect of this adaptation is that it replaces a carboxylation reaction based on HCO_3^- with one based on CO_2 , potentially allowing the new phenotype to thrive in less alkaline environments. It is interesting to note that a similar replacement of pyruvate BC and malate dehydrogenase, between which oxaloacetate is the intermediate, in the opposite rTCA arc is not observed in any organism using rTCA to fix CO_2 . This may simply be because oxaloacetate is the starting precursor to a wide range of anabolic pathways, and is most easily accessed if it is released into solution, and because it is moreover produced as a free intermediate during the cleavage of citrate.

The distribution of the different innovations governing the evolution of the two extant forms of the rTCA cycle may give some insights into the tradeoffs between optimizing growth rate and growth efficiency. The combined fitness advantage of

improving kinetics by fusing the citrate cleavage sequence and improving thermodynamic efficiency through the emergence of the higher-specificity ICDH appears to be significant under most conditions. Only one family (Aquificaceae) within one clade (Aquificales) still uses the ancestral rTCA variant, while the conventional form is distributed across a wide range of bacterial clades. It is further interesting to note that in the vast majority of cases the two innovations occur together. In only one known case (*S. azorense*) has the fusion of the citrate cleavage reaction taken place without elimination of the α -ketoglutarate BC. Characterization of additional Aquificales could help us disentangle the ordering and relative advantage of the two innovations.

Amino Acid Biosynthesis

For most amino acid biosynthetic pathways in *A. aeolicus* the genome annotation leaves little doubt about the correct completion. Alanine, glutamate and aspartate are only one amination reaction removed from intermediates in the rTCA cycle, while an additional amination reaction leads to asparagine (from aspartate) and glutamine (from glutamate). The 3-step sequence from aspartate to homoserine provides the branching point from which the synthesis of threonine, methionine, and lysine (*via* the DAP pathway [62]) diverge. Arginine and proline are both derived from glutamate, with arginine synthesis proceeding *via* ornithine and a partial urea cycle. Histidine is synthesized from ATP and PRPP through the standard histidine synthesis pathway. The shikimate pathway [63] produces chorismate, from which the syntheses of the aromatic amino acids tryptophan, phenylalanine and tyrosine diverge.

Surprising pathway variants used by *A. aeolicus* for the synthesis of several amino acids were identified as a result of gaps in conventional pathways. Subsequent analysis further showed those pathways to represent the likely ancestral pathways to those compounds. In the previous section we mentioned the synthesis of glycine, serine, and cysteine, which in this organism are derived directly from CO_2 and NH_3 (and H_2S for cysteine) through a partial WL pathway coupled to the reductive (= biosynthetic) operation of the “glycine cycle” [2]. Next we discuss the biosynthesis of the branched chain amino acids valine, leucine and isoleucine.

Branched chain amino acids. It has long been thought that most organisms synthesize α -ketobutyrate, a central precursor to isoleucine, by deaminating threonine [64]. However, an increasing number of species have been found to instead derive α -ketobutyrate from pyruvate and acetyl-CoA through what is known as the “citramalate” pathway (see Fig. 4). Originally discovered, and later described in detail, in the Spirochetes [65,66], this pathway was subsequently discovered in a range of species across the bacterial and archaeal domains. It was found to be used in members of both the Euryarcheota [67–70] and Crenarcheota [71,72], as well as Firmicutes (Clostridia) [73,74], Chloroflexi [75], Chlorobia [76], Cyanobacteria [77], and several Proteobacteria [78–80].

Like the observation that the ancestral form of rTCA had an increased degree of symmetry that could have allowed a smaller regulatory structure, the sequence of local functional group transformations in the citramalate pathway is repeated in the synthesis of leucine from α -ketoisovalerate (see Fig. 4). In contrast to the symmetric rTCA variant where parallel reactions are catalyzed by homologous enzymes, however, in this case the parallel reactions are in fact catalyzed by the same enzymes in both pathways [66]. The only exceptions are the pair of acetyl-CoA addition reactions (5, 5') that initiate the two sequences, which are catalyzed by separate, though still homologous enzymes.

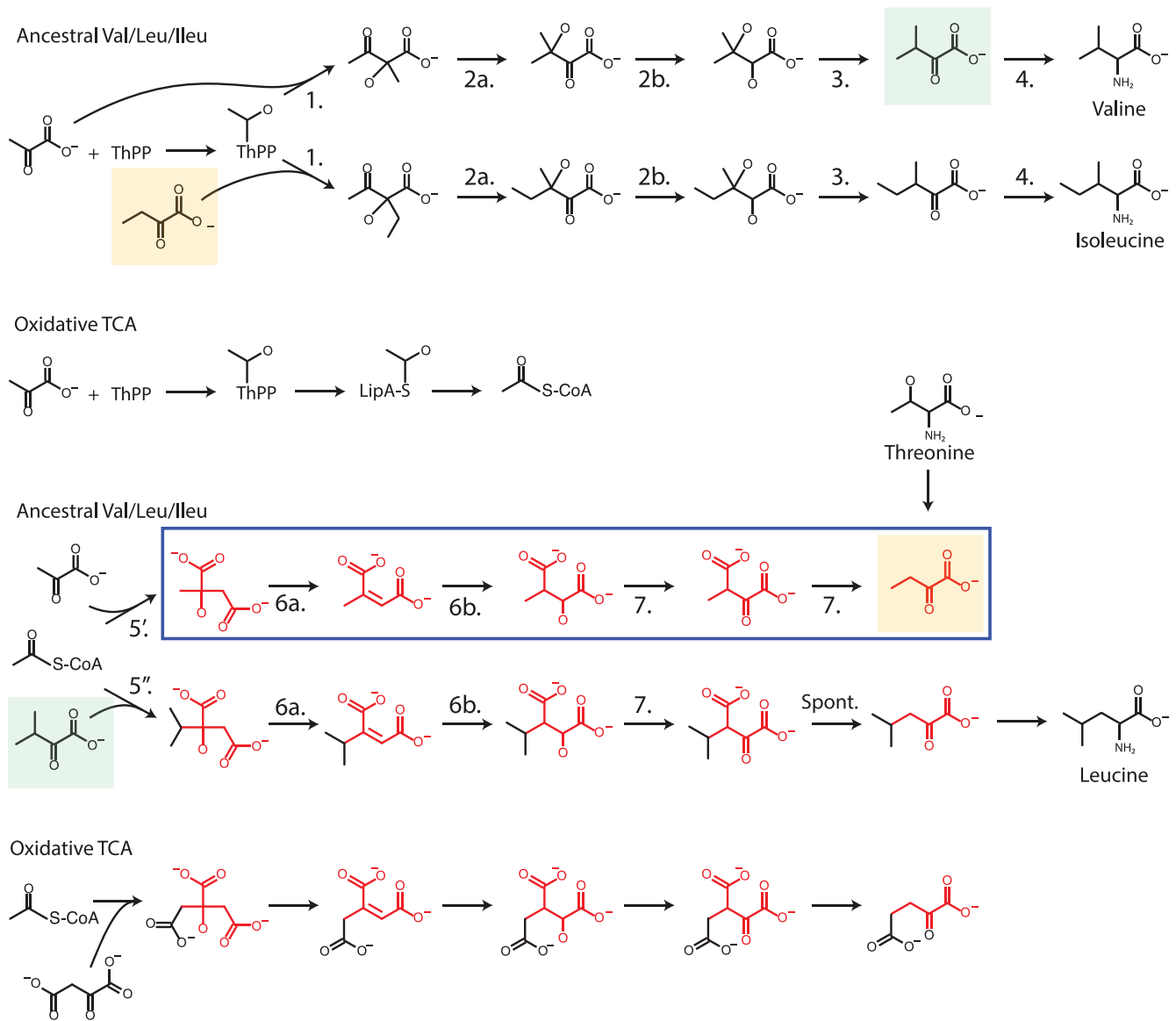


Figure 4. Branched chain amino acid biosynthesis. The chemical sequences show the parallels in terms of local functional group chemistry within the reconstructed ancestral pathways to valine, leucine and isoleucine. The blue box highlight the citramalate pathway of α -ketobutyrate synthesis, reconstructed here to represent the ancestral sequence to this compound. The molecules highlighted in orange and green in turn show the compact interconnectedness of the ancestral pathways to the branched chain amino acids. Parallels to substrate sequences within the oxidative TCA are also highlighted, as well as the alternate route to α -ketobutyrate from threonine.
doi:10.1371/journal.pone.0087950.g004

Together with the broad observed distribution, this leads us to propose that the citramalate pathway represents the ancestral pathway to isoleucine, with threonine deaminase (often called threonine dehydratase, TDH) a more recent innovation.

The genome of *A. aeolicus* is consistent with this hypothesis, as it lacks the gene for threonine deaminase and instead contains two genes annotated as 2-isopropyl malate synthase (IPMS, reaction 5'') [11]. To confirm the presence of the citramalate pathway, and to place this use in evolutionary context, we performed broad genomic surveys for threonine deaminase (TDH) and citramalate synthase (CMS, reaction 5'). Presence of either gene satisfies the constraint of pathway completeness in PMA, as the other genes necessary are shared between both pathways. Distributions of the two genes/pathways are shown in Table 3.

We note briefly that there is some uncertainty in the annotation of CMS, because of the sequence similarity not only to IPMS, but also to homocitrate synthase of the AAA lysine synthesis pathways [70]. However, the pooling of evidence in PMA lowers these uncertainties. For example, if a strain has two genes annotated as IPMS, possesses no gene for TDH, nor any of the other genes of the AAA lysine pathway to which homocitrate synthase might connect, one of the IPMS copies most likely instead codes for CMS. In addition, BLAST search comparisons among these genes were used for cross-validation. For most species with multiple genes annotated as IPMS, BLAST searches show that at the clade level these genes separate into clear groups based on sequence similarity. For most clades, genes within one of these groups were identified in the various experimental studies mentioned above as likely encoding CMS, thus anchoring the group as a whole.

Table 3. Distribution of Isoleucine biosynthesis pathways.

	TDH	CMS	both
Domain/Clade/Family			
Archaea (122)	58	98	47
Crenarcheota (40)	34	30	27
Euryarcheota (79)	21	65	17
Korarcheota (1)	1	1	1
Thaumarcheota (2)	2	2	2
Bacteria total (321)	213	138	55
Bacteria w/o Firmicutes (118)	57	78	29
Aquificales (9)	0	9	0
Thermotogales (13)	5	6	5
Chlorobiales (11)	0	11	0
Chloroflexi (16)	10	12	6
Nitrospirae (3)	0	3	0
Planctomycetes (6)	1	6	1
Verrucomicrobia (4)	1	4	1
Deinococcus-Thermus (15)	10	9	9
Deinococcales (7)	6	1	0
Thermales (8)	4	8	4
Cyanobacteria (41)	30	21	12
Prochlorococcus (12)	12	0	0
Synechococcus (11)	8	5	2
Others (18)	10	16	10
Firmicutes (203)	156	60	26
Bacilli (104)	101	9	9
Clostridiales (99)	55	51	17
Clostridiaceae (33)	26	8	3
Thermoanaerobacter (24)	3	20	1

doi:10.1371/journal.pone.0087950.t003

Combining all these different lines of evidence leads to the distributions shown in Table 3.

Among the archaea in our study, a vast majority (98/122, or 80%) possess a gene for CMS. By contrast, less than half (58/122, or 47.5%) possess a gene for TDH. Moreover, when TDH is found it is mostly co-present with CMS, while in many cases CMS represents the exclusive pathway to isoleucine. This is relevant, because two versions of TDH are known to exist: an 'anabolic' version whose activity is obligatory for cell growth, and a 'catabolic' version that only becomes active during salvage of excess threonine [64]. It has previously been noted that the archaeal TDH best matches the catabolic version of this enzyme [81]. This is consistent with experimental observation that for archaea which have both TDH and CMS, the citramalate pathway represent either a major or the exclusive pathway used in synthesis of isoleucine [68,71]. Thus, we conclude that the citramalate pathway represents the ancestral pathway to isoleucine in archaea, with TDH representing a later innovation that initially emerged for salvage purposes.

Bacteria present a more complex picture than archaea regarding the synthesis of isoleucine. Both TDH and CMS are common in deep-branching bacterial clades, with TDH occurring with higher frequency. However, this balance is dominated by Firmicutes, and among other deep-branching bacteria, the

citramalate pathway appears to be the most abundant route to isoleucine. Strikingly, a significant number of deep-branching clades whose members include (hyper)thermophiles, autotrophs, or both, show near exclusive use of the citramalate pathway: Aquificales, Chlorobiales, Nitrospirae, Planctomycetes, and Verrucomicrobia.

A closer look within clades where TDH represents the majority pathway provides further context for evolutionary reconstruction. For example, within the large and diverse Firmicute phylum, the aerobic Bacillales nearly all possess the threonine pathway, while the anaerobic Clostridiales exhibit the threonine and citramalate pathways in about equal frequencies. Within the Clostridiales in turn, the Clostridiaceae family, which contains many pathogenic strains [82], make nearly exclusive use of the threonine pathway, while the Thermoanaerobacteriales family, which like the Aquificales are hyperthermophiles, make nearly exclusive use of the citramalate pathway.

Within the Cyanobacteria in turn, the majority of strains using the threonine pathway belong to the genera *Synechococcus* and *Prochlorococcus*. Many of the former and all of the latter are divergent cyanobacterial lineages highly adapted to oligotrophic regions of the world's oceans, often dominating those environments [83–85]. Among all other cyanobacteria, the citramalate pathway represents the major route to isoleucine. Finally, within Deinococcales-Thermales [86], which have approximately equal numbers of both pathways, the Deinococcales use mainly the threonine pathway, while the hyperthermophilic Thermales use mainly the citramalate pathway.

Lastly, it has previously been observed that many deep-branching bacteria that possess both TDH and CMS, appear to possess the catabolic version of TDH [81]. In our sample of Thermotogales and Chloroflexi, both of which show frequent occurrence of TDH, BLAST searches show that many of these genes are better matched to the catabolic than the anabolic TDH of *E. coli*. Thus, all evidence combined leads us to conclude that the citramalate pathway is the ancestral pathway also in bacteria, and thus for all life, with the threonine pathway initially emerging as a salvage pathway.

This conclusion is important relative to the previous discussion of a minimal ancestral rTCA cycle. In addition to the enzymes shared between the citramalate pathway and the final sequence in leucine synthesis (reactions 5–7), the enzymes catalyzing the homologous sequences in valine and isoleucine synthesis (reactions 1–3) are similarly shared across pathways, while the final amination reaction (4) is in all three pathways performed by the same enzyme [87]. While some organisms have additional copies of the first thiamin catalyzed reaction (1) for purposes of regulation [88,89], *A. aeolicus* possesses only a single, two-subunit enzyme for both reactions [11]. As mentioned, the lone exception to this pattern of promiscuous catalysis is the acetyl-CoA addition reaction (5, 5') initiating both sequences in leucine/isoleucine synthesis. However, if these reaction had not been mediated by thioester intermediates, they would represent carboxyl reduction and carbon-carbon bond forming reactions that in carbon-fixation pathways are associated with ATP hydrolysis [55]. Similar to those more constrained carbon-fixation reactions, the high sequence similarity of the two enzymes caused them to be originally annotated as the same enzyme [11,66]. The more complex, constrained nature of these reactions may thus have increased the kinetic advantage of duplication and divergence toward greater enzyme specificity. Thus, if as before we assume an ancestral enzyme with broader substrate affinity for reactions 5/5', then all 21 total reactions in the synthesis of valine, leucine and isoleucine

starting from pyruvate and acetyl-CoA would have required only 7 enzymes for catalysis (see also Fig. 4).

In contrast to the evolution of the rTCA cycle, most enzymes catalyzing homologous reaction sequences in this case have not duplicated and diverged, except for CMS and IMPS. Why this difference? The difference in mass flux between the two pathways would appear to offer a straightforward explanation. While most cellular carbon passes through the rTCA cycle in autotrophs using that pathway, only 3 out of 20 amino acids are generated through the pathways in Fig. 4. Thus, even if we simplistically assume roughly equal biomass partitioning between amino acids, nucleotides and lipids, the mass flux difference between the two sets of pathways is about an order of magnitude. The only enzyme for which selection pressure to improve kinetics appears to have led to duplication and divergence within the synthesis of branched chain amino acids represents a more complex, possible thermodynamic bottleneck reaction.

Emergence, and combined regulation with the rTCA cycle. Additional overlaps between the rTCA cycle and the branched chain amino acid biosynthetic pathways may provide insights into how the latter emerged. The dehydration/hydration isomerization and dehydrogenation sequences (reactions 6a, 6b, 7) in the reconstructed ancestral pathways to leucine and isoleucine is homologous to the sequence of local group transformations occurring in the opposite direction in the large-molecule half of rTCA (reactions 4', 5' and 5* in Fig. 2). The associated enzymes in the two sub-systems may also have a common origin. In *A. aeolicus* (which uses the ancestral rTCA variant), the dehydration/hydration-isomerization reaction is catalyzed by a single subunit aconitase enzyme (ACO) in rTCA, and a two subunit enzyme (LeuC/D) in the branched chain amino acid pathways. However, the combined length of LeuC and LeuD is similar to that of ACO, while LeuC and LeuD also have high sequence similarity to consecutive, adjacent portions within ACO (data not shown). This suggests the possibility that following duplication and divergence from a common ancestral enzyme the two subunits became fused within the rTCA cycle but not the branched chain amino acid pathways, due to the differences in mass flux density. Similarly, the homologous (de)hydrogenation reactions in the two sub-systems are in *A. aeolicus* also catalyzed by enzymes with high sequence similarity. The enzyme homologies do not extend to the (de)carboxylation reactions across the two sub-systems, which in the direction of decarboxylation is facile, and in the direction of carboxylation requires complex cofactors and ATP hydrolysis (see previous discussions). Phylogenetic reconstructions of the lineages of these enzymes could shed additional light on these hypotheses.

These observations may thus suggest that the existence of a sequence of substrate chemistry operating in the opposite direction within the ancestral rTCA cycle could have facilitated the emergence of the ancestral pathways to leucine and isoleucine. The only truly new form of chemistry in the citramalate pathway and its homologous sequence in leucine synthesis is the initiating reaction involving ligation of acetyl-CoA (reactions 5, 5' in Fig. 4). Moreover, the facile decarboxylation of β -carboxylic acids, which in the reductive direction of the rTCA cycle may have increased the cost of an already complex reaction, may have created an advantage when used in the opposite direction, possibly further facilitating the emergence of the ancestral citramalate/leucine sequence.

The pattern of pathway diversification by innovation of the initiating reaction, followed by re-use of similar or identical downstream enzymes to catalyze homologous reaction sequences, was also the primary mode of diversification proposed for carbon-fixation pathways in Ref. [1]. The difference to those previous

proposals is that here also the direction of the re-used sequence chemistry is suggested to have changed. We should also note that in comparing these innovations some caution should be exercised because the establishment of the first pathway to a set of amino acids is an innovation occurring in the era prior to LUCA, while diversification of carbon-fixation pathways occurred after LUCA. Nonetheless the parallels in the suggested modes of innovations is striking and may suggest a general principle of early metabolic evolution.

If true, this hypothesis of pathway evolution may have allowed the LUCA to regulate the combined sub-systems with an even smaller genome than we have suggested for them separately. If the homologous enzymes in the two sub-systems arose through duplication and divergence from a common ancestor, then the two sequences could have potentially been catalyzed by the same enzymes in an earlier era. We suggested above that the complete ancestral forms of rTCA and branched chain amino acid biosynthesis could in an era of more promiscuous enzymes each have been catalyzed by only 7 enzymes total. The added observations here suggest that the entire connected network of rTCA plus the pathways to the branched chain amino acids could have been catalyzed by only 12 total enzymes.

Relation to the reversal of the TCA cycle. The key innovation of the acetyl-CoA ligation reaction that initiates the homologous sequences within leucine and isoleucine synthesis, and possibly governed their emergence, may have also facilitated the later emergence of other pathways within metabolism, in particular the reversal of direction of the TCA cycle. The entire homologous reaction sequences (5–7) in the citramalate and leucine pathways are also known as “keto acid elongation” sequences [19], and are further repeated within the oxidative TCA cycle (as well as the AAA lysine synthesis pathway). It is becoming clear that the TCA cycle originally operated in the reductive direction (see Ref. [1] and references therein for discussion), which means that keto acid elongation likely occurred within branched chain amino acid synthesis prior to its use in the oxidative TCA cycle. The thiamin-facilitated decarboxylation of pyruvate (see Fig. 4) thus similarly appears to have been used first in the synthesis of branched chain amino acids then in its use within the oxidative TCA cycle. Finally, lipoic acid is used only in the oxidative (and not the reductive) direction of the TCA cycle, while its likely ancestral function was in the synthesis of glycine through the glycine cycle [2] (see also the section on lipoic acid below for additional supporting evidence).

Thus, we suggest that at the substrate level the prior existence of the biosynthetic pathways leading to valine/leucine/isoleucine facilitated the reversal of the TCA cycle from the reductive to the oxidative direction. The additional key innovation appears to have been the recruitment of lipoic acid from the glycine cycle to its interaction with thiamin in the production of acetyl-CoA from pyruvate. Broad affinity of earlier enzymes would have aided the emergence of this novel pathway as promiscuous activity could have allowed this pathway to proceed at lower rates, with duplication and divergence later being favored as respiration came fully online and mass flux through this pathway increased.

Cofactors

Cofactors are a distinct class of molecules at the substrate level of metabolism, forming a chimeric intermediary layer between monomers and polymers in terms of structure [90]. Cofactors are also critical components of the control hierarchy of metabolism, facilitating many key reaction mechanisms, and thus the overall integration of metabolism. Each cofactor generally facilitates a distinct and specialized catalytic function, and their emergence can

thus be thought of as the outgrowth of kinetic feedback loops, each “opening up” new spaces in the universe of organic chemistry and bringing them under the control of biology [1]. Understanding the evolution of cofactor biosynthesis is thus important both in providing context to discussions on the origin of life, as well as understanding major physiological lineages in the tree of life. In this section we focus on the synthesis of several cofactors in *A. aeolicus*, using the reconstruction to provide additional insights into the evolution of their functions.

Lipoic acid. Lipoic acid is a cofactor with very limited, but key metabolic roles. Lipoic acid is central to the “Glycine Cleavage System” (GCS) [91], which connects glycine and serine metabolism to folate one-carbon chemistry, and is also used (as previously mentioned) in the ferredoxin:oxidoreductase decarboxylation reactions in the oxidative TCA cycle and the degradation of branched chain amino acids. The GCS is known to be reversible [91], has nearly neutral thermodynamics [55], and likely originally operated in the reductive (i.e. biosynthetic, not degradative) direction as part of the ancestral pathway leading to glycine and serine [2]. For these reasons the GCS together with serine methyl transferase (SMT) has also been called the “Glycine Cycle” [2]. The phylometabolic analysis that places reductive glycine synthesis at the base of the tree of life, as part of the phenotype of the last common ancestor, suggests that the function of lipoic acid in the glycine cycle preceded its use in either the oxidative TCA cycle or the degradation of branched chain amino acids. This provides important context for interpreting the distribution of lipoic acid biosynthesis genes that we discuss next.

Three pathways are known for the biosynthesis of lipoic acid (see Fig. 5). The conventional pathway, first described in *E. coli* [92], involves transfer of an octanoyl moiety from the Acyl Carrier Protein (ACP) to one of the Lipoyl Dependent (LD) enzymes, followed by sequential sulfuration of the octanoyl moiety to produce the final lipoated enzyme [93]. The first step is catalyzed by octanoyl transferase (LipB), while the second is catalyzed by Lipoyl Synthase (LipA). A variant of this scheme was recently discovered in *B. subtilis*, which involves the same basic chemistry, but distinct set of enzymes and an additional intermediate in the transfer of octanoate. In *B. subtilis*, the distinct octanoyl transferase LipM transfers octanoate from ACP to the H-protein of the GCS, followed by a second transfer (catalyzed by LipL) to the E2 subunit of pyruvate dehydrogenase [94–96]. Both LipM and LipL had previously been obscured due to their sequence similarities to LplA of *E. coli* [95]. A third distinct pathway was recently discovered in an *E. coli* mutant in which LipB had been deactivated. In this mutant, lipoate protein ligase (LplA), normally used in the attachment of free lipoic acid to LD enzymes, is recruited in the transfer of free octanoate to an LD enzyme through an AMP-bound intermediate [93,97–99]. In light of these variations, it is noteworthy that *A. aeolicus* lacks a gene for LipB and is annotated as having LipA and LplA [11]. This raises the question, which of the pathway alternates does *A. aeolicus* use, and what does this teach us about how the biosynthetic pathways and the functionality of lipoic acid evolved?

To examine this question, we performed broad genomic surveys for each of the enzymes used in lipoic acid metabolism, shown in Table 4. In addition to LipA, LipB, LplA, LipM/LipL, our survey also includes LplJ, a distinct lipoate protein ligase found in *B. subtilis* [95]. It can immediately be seen that the conventional LipA+LipB combination is not widely distributed across deep-branching clades. Only 6 archaeal strains, nearly all (5/6) in the Thermoproteale family within the Crenarcheota, appear to possess this pathway. Among deep-branching bacteria, only the Deinococcales, Cyanobacteria and Clostridiales family within Firmicutes

show significant abundance of the LipA+LipB combination. In contrast, LplA (and its combination with LipA) is widely and more evenly distributed across both archaea and bacteria. However, BLAST searches indicate that in most of these cases LplA is in fact a better match to LipM of *B. subtilis* than to LplA of either *E. coli* or the euryarchaeota *T. acidophilum* [100]. While this suggests that the *B. subtilis* variant may be the ancestral pathway to lipoic acid, in most cases we could not find a LipL gene to accompany the putative LipM gene, presenting a puzzle as this gene is absolutely required in lipoic acid synthesis in *B. subtilis* [95].

This is where the functional roles of lipoic acid provides critical evolutionary context. As explained, the likely ancestral function of lipoic acid is its role in connecting glycine/serine metabolism to folate-C₁ chemistry through the glycine cycle, for which it remains (nearly universally) essential to this day. In contrast, the role of lipoic acid in the oxidative TCA cycle or the degradation of branched chain amino acids likely arose later, and is not essential to many organisms. For example, many autotrophs, including *A. aeolicus*, do not use the oxidative TCA cycle nor do they degrade branched chain amino acids. Such organisms should thus not need LipL to transfer octanoate from the H-protein of GCS to other LD enzymes, because they do not possess them. Instead, direct sulfuration of octanoate bound to the H-protein of GCS, a reaction known to be catalyzed by LipA [101], is sufficient to allow the sole function of lipoic acid in these organisms. This proposed sequence of enzyme functions is supported by the observation that in deep-branching clades that do not possess the glycine cycle – including many Clostridia, all methanogenic families within the Euryarcheota or the Desulfurobacteriaceae family within Aquificales [2] – we do not find any genes associated with lipoic acid biosynthesis or uptake. Thus, we suggest that in many cases where a putative LipM is found but LipL is absent, a pathway is used in which LipM is directly followed by LipA (see Fig. 5), and we also propose that this represents the ancestral pathway for *de novo* lipoic acid biosynthesis. Note that this proposal further reinforces the conclusion that the reductive TCA cycle preceded the oxidative version.

In this scenario, an additional octanoyl transferase (LipL) emerged for the second subsequent transfer to the E2 domain of PDH, possibly through duplication and divergence of LipM as seen in *B. subtilis*. This would have introduced a redundancy into the lipoic acid system by producing two dedicated octanoyl transferase enzymes. This redundancy then appears to have been removed by replacing the two distinct transferase enzymes with a single all-purpose LipB transferase, as seen in *E. coli*. The high sequence similarity among LipM, LipL, LplA and LplJ further suggests that the environmental uptake genes likewise arose through duplication and divergence from octanoyl transferase genes, but that some of this ancestral function was maintained in LplA, making possible its recruitment in *E. coli* mutants lacking LipB.

Vitamin B6. In contrast to the narrow functionality of lipoic acid, vitamin B6, which refers to pyridoxal 5-phosphate and its substitutes, is one of the most functionally diverse cofactors, with its different forms facilitating a very wide range of reaction classes [102]. Its diverse functionality and relatively simple chemistry, plausibly accessible under abiotic conditions, has led to the suggestion that it may have been one of the earliest cofactors [103,104].

There are two known biosynthetic pathways leading to vitamin B6 in modern metabolism. In the first recognized pathway, described in *E. coli*, pyridoxine phosphate is derived from 4-erythrophosphate [102]. This pathway is known as the ‘DXP dependent’ pathway, because 1-deoxy-D-xylulose-5-phosphate

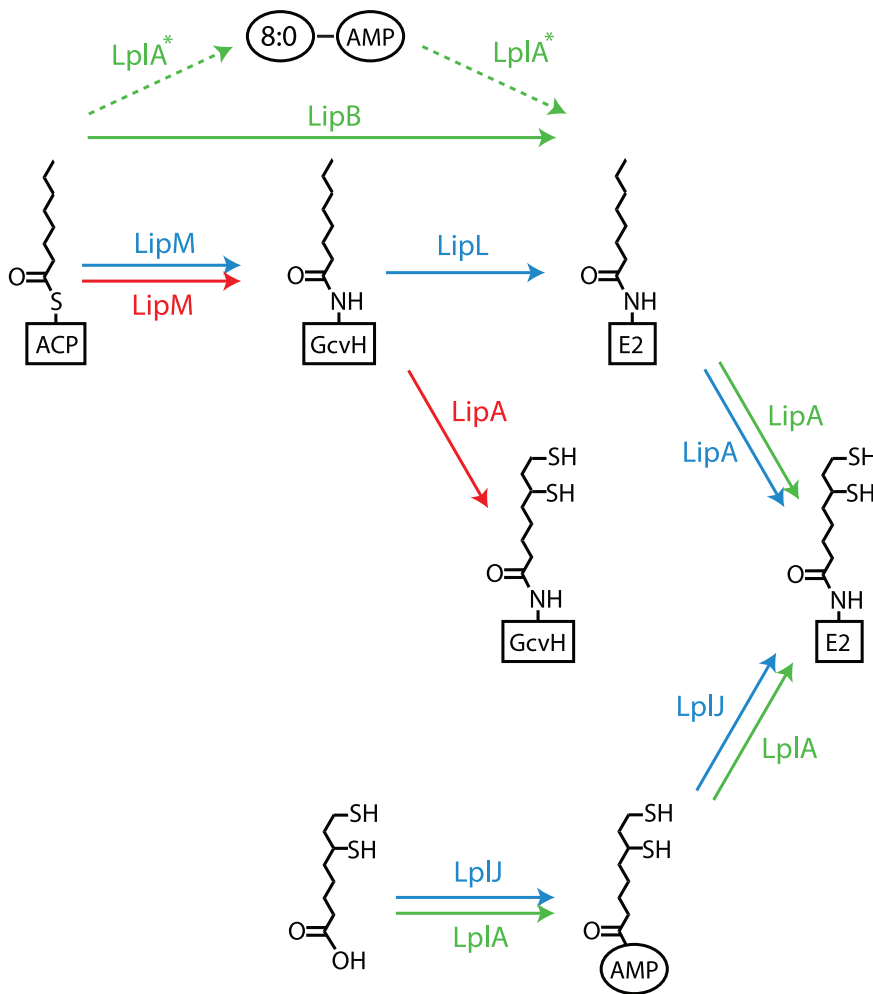


Figure 5. Lipoyl acid biosynthesis and lipoyl-protein assembly. In *E. coli* (green sequence), octanoate is transferred from ACP to the E2 subunit of pyruvate dehydrogenase (PDH) by LipB, followed by sulfuration to lipoic acid by LipA. In *E. coli* mutants lacking LipB, octanoate is transferred through an alternate route with an AMP-bound intermediate by LplA, normally used for incorporation of free lipoic acid. In *B. subtilis* (blue sequence), octanoate is transferred from ACP first to the H-protein of GCS by LipM, followed by a second transfer to the E2 subunit of PDH by LipL. *B. subtilis* also uses LplJ instead of LplA for incorporation of free lipoic acid. In red is the suggested ancestral biosynthesis of lipoic acid (see main text). doi:10.1371/journal.pone.0087950.g005

(DXP) is the secondary input to the final condensation reaction that produces the pyridine ring of pyridoxine. In the second (DXP independent) pathway, first described in *B. subtilis*, pyridoxal phosphate is synthesized through the direct condensation of ribulose phosphate and glyceraldehyde phosphate [105,106].

Previous analyses found genes for the DXP-independent pathway to be highly conserved and distributed across both archaeal and bacterial domains, while genes for the DXP-dependent pathway were found mainly in the γ -proteobacteria, suggesting this latter pathway emerged later in evolution [107,108]. However, *A. aeolicus* possesses the DXP-dependent pathway, prompting us to further examine this hypothesis. Table 5 shows the distribution of the key enzymes involved in the condensation steps in both sequences – PdxA/J in the DXP-dependent pathway and PdxS/T in the DXP-independent pathway – across bacteria and archaea.

These distributions show some striking patterns. It had previously been noted that the two pathways are mutually exclusive within organisms, which use only one or the other [107]. Our analysis shows that the pathways are in fact mutually exclusive at the *clade* level, much more so than we have seen for

pathway variants in other sub-systems we have previously analyzed. PdxA is found in a few species within both archaea and Firmicutes, but that enzyme catalyzes a hydrogenation reaction, with the enzyme catalyzing the actual ring condensation reaction (PdxJ) completely absent in those cases. Our analysis further appears to confirm that the DXP-independent pathway represents the ancestral pathway. In addition to nearly all archaea, several deep-branching bacterial clades (Thermotogales, Chloroflexi, Deinococcales, Firmicutes) also use this pathway.

It was previously suggested that the DXP-dependent pathway arose within proteobacteria [108], but the fact that all members of several deep-branching clades use this pathway suggests it may actually have been an earlier innovation. The observed distribution of both pathways is difficult to explain, however. That the pathways are mutually exclusive at the clade level requires either early HGT to progenitors of clades, extensive gene transfer between select clades after they had diverged, or extensive transfer within clades that can take the appearance of genes sweeping through the population [17]. In any of these cases there is no obvious explanation for why transfer would have been restricted to

Table 4. Distribution of lipoic acid biosynthesis genes.

	LipA	LipB	A+B	LpIA	A+A	LipM	LipL	A+L+M	LpIJ
Domain/Clade/Family									
Archaea total (126)	37	6	6	60	31				
Crenarcheota (41)	18	5	5	27*	13				
Euryarcheota (80)	19	1	1	32	18				
Korarcheota (1)	0	0	0	1*	0				
Nanoarchaea (1)	0	0	0	0	0				
Thaumarcheota (3)	0	0	0	0	0				
Bacteria total (329)	214	105	102	100	83	94	108	85	146
Aquificales total (9)	7	3	3	4*	4				
<i>Aquificaceae</i> (4)	4	0	0	4*	4				
<i>Hydrogenobacter</i> (3)	3	3	3	0	0				
<i>Desulfotobacter</i> (2)	0	0	0	0	0				
Thermotogales (13)	0	0	0	13*	0				
Deinococcus-Thermales (15)	15	15	15	2	2				
Chloroflexi (16)	8	6	6	10*	8				
Chlorobi (11)	11	1	1	11*	11				
Planctomycetes (6)	5	1	0	5*	5				
Nitrospira (3)	2	0	0	3*	2				
Verrucomicrobia (4)	4	3	3	4*	4				
Cyanobacteria (44)	43	43	43	43*	43				
Firmicutes total (208)	119	33	31	5	4	94	108	85	146
<i>Bacilli</i> (106)	92	7	7	0	0	85	106	85	96
<i>Clostridia</i> (102)	27	26	24	5	4	9	2	0	50

Within the column counting LpIA genes, entries denoted by (*) are often better matched to LipM of *B. subtilis* than LpIA of either *E. coli* or *T. acidophilum*. See main text for additional details.

doi:10.1371/journal.pone.0087950.t004

occurring only between select clades, nor is a selective advantage apparent.

An alternative explanation would be (possibly repeated) convergent evolution early in the divergences of clades. This explanation has some appeal, as the key enzymes in both pathways, PdxA/J and PdxS/T, are in fact very similar in their 3-D structure and in the sequence of local functional group transformations that make up the respective condensation reactions [102]. However, even for convergent evolution we are lacking a good explanation for why only some clades would pervasively adopt this new strategy, while others did not.

If the evolutionary sequences and driving forces are not clear, we can at least identify features of the two pathways that would have facilitated the transition between them. The shared fold structure and similarity in reaction mechanisms, but low sequence similarity, between PdxA/J and PdxS/T have been interpreted to mean that they represent convergent discoveries [102]. However, it is also possible that PdxA/J emerged from PdxS/T and that both have been under strong selection pressure, causing their sequence to diverge strongly.

Another common feature between pathways is they both start from intermediates within the pentose phosphate pathway. The key exception is DXP itself, which is the product of the first committed reaction in the DOXP pathway of terpenoid backbone synthesis. Archaea exclusively use the alternate mevalonate (MVA) pathway to synthesize terpenoids, while bacteria use both the MVA and DOXP pathways [109], possibly providing a partial

explanation for why the DXP-dependent pathway to vitamin B6 emerged only within bacteria and not archaea.

Finally, the other reactions in the DXP-dependent pathway that lead up to the condensation sequence catalyzed by PdxA/J represent common and widely used metabolic reactions catalyzed by members of highly diversified enzyme families. The reaction sequence connecting 4-erythrophosphate to phospho-4-hydroxy-threonine, the input to the PdxA/J-catalyzed ring condensation sequence, consists of a hydration/reduction of an aldehyde to a carboxyl group, a subsequent dehydrogenation of an alcohol to a carbonyl group, and finally the reductive amination of that carbonyl group. Especially in an earlier era of more promiscuous catalysts, this pathway could thus well have been recruited *en bloc* into the emergent DXP-dependent pathway. The selective advantage of this adaptation, and the way it might have led to the peculiar distribution of both pathways, remains to be explained, however.

Quinones. The main component of the quinone pool in *A. aeolicus* was determined to be 2-demethylmenaquinone-7 (DMK-7) [110]. Other Aquificales, including its close relative *H. thermophilus*, had previously been found to use 2-methylthiomenaquinone-7 [111–113]. Menaquinone (MK) has significantly lower redox potential than ubiquinone (UQ), and, based on distributions of these two quinone types both across the tree of life [114,115] and within clades known to bridge the anaerobic-aerobic domains, UQ was suggested to have emerged with the rise of atmospheric oxygen [116]. Membrane-dissolved quinones exchange electrons directly with the fumarate/succinate redox couple, which is

respectively an electron acceptor or donor depending on the direction of this reaction [114,117]. The possible emergence of the higher potential UQ with the rise of oxygen may thus have allowed the fumarate-succinate interconversion to reverse to the oxidative direction, further facilitating reversal of the TCA cycle as a whole. Generally, whereas reduced MK is easily oxidized in the presence of oxygen, disrupting electron flow into biosynthesis, reduced UQ is stable in the presence of oxygen [116]. The redox potential of the de-methylated version of menaquinone, DMK, lies between that of MK and UQ, possibly reflecting the microaerophilic character of *A. aeolicus* [110].

Nucleotides

Assignment of biosynthetic pathways to pyrimidines and purines was mostly unambiguous in *A. aeolicus*. At the substrate level there is little major variation in the biosynthesis of these compounds [118], and the genome of *A. aeolicus* shows complete gene sets for their synthesis [11]. There is some ambiguity in the interconversion between differently substituted purines and pyrimidines due to the well-known broad substrate affinity of many of the enzymes involved (e.g. [119,120]). We therefore did not significantly modify the conservative broad assignments made by SEED in this sub-network. Experimental studies would probably be needed to elucidate the fine scale activity/regulation of these reactions if it were deemed important for a highly quantitative model. We again suggest that the broad affinity of enzymes interconverting differently substituted nucleobases indicates that the lower mass flux of these reactions (for example compared to the rTCA cycle) significantly reduces the benefit relative to the cost of using multiple more-specific enzymes.

There are a few noteworthy details in the biosynthetic pathways of nucleotides. The initial synthesis of the IMP backbone involves several steps in which formyl groups are incorporated, which can proceed either through an ATP mediated addition of free formate, or through donation of the formyl group by N¹⁰-Formyl THF [118]. Archaea that possess tetrahydromethanopterin (H₄MPT) rather than tetrahydrofolate (THF) as their C1 carrier use free

formate in purine synthesis because H₄MPT is not a good donor of formyl groups [43,121]. Most other organisms use THF to transfer formyl groups in purine synthesis [118], while *E. coli* was found to possess both mechanisms [122]. *A. aeolicus* follows these trends and uses THF as the formyl donor during purine synthesis.

Another minor variation in the synthesis of purines involves the carboxylation of aminodazole ribonucleotide (AIR). Like other bacteria, *A. aeolicus* uses a 2-step incorporation of HCO₃⁻ involving ATP hydrolysis for this reaction. By contrast, higher organisms use a 1-step incorporation of CO₂ in an ATP free system, in which the enzyme is moreover often fused to the enzyme catalyzing the subsequent reaction [118]. Similar to enzyme replacements we have seen in other metabolic sub-systems, this may represent another adaptation selected because it improves both thermodynamic efficiency and pathway throughput, in this case also shifting dependence from HCO₃⁻ to CO₂ as a secondary effect.

Pyrimidine synthesis in *A. aeolicus* again reflects the primitive nature of its metabolism. In the first reaction in this pathway carbamoyl phosphate is synthesized from glutamine, HCO₃⁻, and ATP. Experimental studies showed that in *A. aeolicus* this three-part reaction is catalyzed by a heterotrimer enzyme with relatively inefficient coupling between the subunits [123]. By contrast, in *E. coli* two of those subunits are fused together, resulting in a heterodimer enzyme, while in mammals all three subunits plus the enzymes for the subsequent reactions to carbamoyl-aspartate and dihydroorotate are fused together into one large single subunit enzyme [124]. Paralleling the suggested ancestry of ATP citrate lyase, the collection of observations about pyrimidine synthesis suggest that the heterotrimer carbamoyl-phosphate synthetase of *A. aeolicus* is more closely related to the ancestral enzyme for this reaction [123,124]. However, while the cost of improving kinetics through gene fusion is lower than other cases of duplication and divergence, the lower mass flux density of pyrimidine synthesis relative to core carbon fixation also reduces the benefit of fusion. This may help explain why these genes are also not fused in many other bacteria. For *A. aeolicus* another reason that fusion did not take place may be that the heterotrimer structure may provide additional stability under hyperthermophilic conditions [124].

Table 5. Distribution of pyroxidial phosphate synthesis genes.

Domain/Clade/ Family	Pdx					
	S	T	S+T	A	J	A+J
Archaea total (126)	114	114	113	6	0	0
Bacteria total (339)	226	206	206	123	77	76
Aquificales (9)	0	0	0	8	9	8
Thermotogales (13)	12	12	12	0	0	0
Chloroflexi (16)	16	15	15	0	0	0
Planctomycetes (6)	0	0	0	6	6	6
Nitrospirae (3)	0	0	0	3	3	3
Chlorobia (11)	0	0	0	11	11	11
Deinococci (15)	15	15	15	0	0	0
Verrucomicrobia (4)	0	0	0	4	4	4
Cyanobacteria (44)	0	0	0	44	44	44
Firmicutes total (208)	183	164	164	47	0	0
Bacillales (106)	106	106	106	11	0	0
Clostridia (102)	77	58	58	36	0	0

doi:10.1371/journal.pone.0087950.t005

Cellular Encapsulation

The cellular encapsulation of *A. aeolicus* consists of three main components: phospholipid membranes, a peptidoglycan cell wall, and lipopolysaccharide. *A. aeolicus* has the full gene complement for the standard diamminopimelate-based variant of peptidoglycan synthesis that is common to gram-negative bacteria [125], but leaves substantial gaps within lipopolysaccharide synthesis pathways. These pathways remain an important area of experimental study, as they have required a large number of gap-fills for which we lacked overall context in the curation process. Of the three components of encapsulation, the composition of phospholipids contains the most information on the ecology of *A. aeolicus*.

Lipid biosynthesis. Lipid metabolism represents the single largest sub-system within the reconstructed network of *A. aeolicus*, containing nearly 200 out of ~760 reactions. As mentioned this is partly due to the fact that we explicitly represent each reaction within this sub-system, and partly due to the fact that *A. aeolicus* has a complex and diverse lipid composition (see Supplementary Table I) [126]. However, the elongation of all fatty acid chains is a polymerization sequence in which 2-carbon units (from acetyl-CoA) are added, and then reduced, through a repeated sequence of the same 7 reactions catalyzed by the same 8 enzymes, with only the length of the fatty acid tail away from the reaction site varying [38]. Much of the size of this sub-network thus reflects

representation in the model rather than the associated genome content.

In addition, the fatty acid sub-network is further expanded due to the fact that *A. aeolicus* uses fatty acid chains that contain methyl groups, propyl rings, and unsaturated bonds at different positions [126]. Each of these different substitutions is introduced at a different point during the elongation process, resulting in an expanding set of intermediates that is tracked in the network prior to the output of chains of different lengths and substitutions into the final lipid assembly process.

Finally, the lipid content of *A. aeolicus* is unusual among bacteria for containing both phospho-ester and phospho-ether lipids [126]. In general, most bacteria use fatty acid-based ester lipids, while archaea use isoprenoid-based ether lipids [127]. The additional use of fatty acid ether lipids by *A. aeolicus* thus represent a sort of intermediary strategy.

Altogether lipid biosynthesis can be thought of as a compact and highly modular system that distributes 2-carbon units over a set of states of different lengths and substitution patterns that can be varied depending on environmental context. Methyl group side-chains, unsaturated bonds or cyclopropane rings can be used to modify the fluidity of the membrane, while cyclopropane rings may also be used to adapt to lower pH [128]. The linkage of fatty acids to the the glycerol backbone can in turn be varied between ether or ester bonds to modify the permeability of the membrane [129]. Apart from basic inputs and final assembly (isoprenoids vs. fatty acids, ethers vs. esters), the regulation of lengths and substitution patterns appears to be the main factors permitting wide variability in lipid composition. The diverse and varied composition of *A. aeolicus* lipids, including both ester and ether linkages, appears to reflect the “stressed” hyperthermophilic conditions of the hydrothermal vents and springs where it lives.

Energy Metabolism

Energy metabolism has the highest mass flux density of all cellular processes, because it generates the global energetic driving forces (both reductants and ATP) for all subsequent metabolic interconversion. The energy metabolism of *A. aeolicus* represents one of its most studied aspects [5], and appears to consist of many tightly interconnected and optimized pathways. Many respiratory proteins are organized in polycistronic operons in the genome, and *in vitro* experiments suggest they are assembled into super-complexes once functionally expressed [5,130,131].

It is possible that these observations may reflect selection for improved kinetics. In contrast to heterotrophs, which can obtain energy from organics and may use fermentative metabolic modes, autotrophs obtain all energy from inorganic sources. The free energy density available from inorganic redox couples used by chemoautotrophs may further be as much as an order of magnitude lower than that provided by sunlight used by photoautotrophs. Together these effects could create a significant selective advantage for improving the kinetics of energy metabolism in chemoautotrophs.

However, there appears to be some debate as to the functional relevance of protein super-complexes within respiratory electron transport chains (ETC). Recent studies of the yeast mitochondrial ETC concluded that protein complexes found to be part of super-complexes in *in vitro* studies, behaved as freely diffusing entities in intact cells [132]. It thus also remains possible that the properties of the super-complexes of *A. aeolicus* energy metabolism primarily reflect adaptation to growth at higher temperature. While interactions between components of other known respiratory super-complexes have generally been found to be rather weak, in *A. aeolicus* super-complexes are found to be exceptionally stable

[5,130,131]. In any case it is clear that the cellular organization, functional integration and operation of *A. aeolicus* energy metabolism remains an important and exciting area of research. Because of their complex interconnected nature, the electron transport chains in this organism represent a potential treasure trove for further insights into the ecology and evolution of extremophilic and/or early life forms. We next briefly summarize some of the main findings to date.

A. aeolicus has a versatile and diverse energy metabolism. Molecular hydrogen is the universal (and obligate) electron donor, but can in some cases be supplemented by hydrogen sulfide (H_2) [133], and possibly elemental sulfur (S^0) or thiosulfate ($\text{S}_2\text{O}_3^{2-}$) [5,134]. An observation that H_2S cannot replace H_2 as the sole electron donor has been argued to be the consequence of tight coupling between respiratory super-complexes, which may prevent uptake of intermediates in the respiratory sequence [131,135].

A variety of compounds can act as electron acceptors. Molecular oxygen (O_2) is the major electron acceptor, and *A. aeolicus* is generally described as a microaerophile [11], which under conditions so far tested has held true. However, whether oxygen is truly obligatory as an electron acceptor in this organism remains unknown, and is challenging to study due to its extremophile lifestyle, interconnected energetic pathways, and resulting dependence on multiple inorganic compounds [5]. For example, the metabolic network indicates that nitrate (NO_3^-) could potentially also be used as a terminal electron acceptor [11], a known ability in many other Aquificales [16,136]. This ability has so far not been reported in *A. aeolicus* [3,5], and the nitrate to ammonia metabolic sequence has instead been suggested to serve mainly a biosynthetic role [11].

Sulfur has a dynamic and varied role in the energetics of *A. aeolicus*, likely in part because of its ability to exist in a wide range of oxidation states [137]. In addition to acting as electron donor (H_2S , possibly S^0 and $\text{S}_2\text{O}_3^{2-}$), several sulfur compounds are capable of acting as electron acceptors. Elemental sulfur and tetrathionate ($\text{S}_4\text{O}_6^{2-}$) act as electron acceptors at the hydrogenase I complex [138]. Thiosulfate ($\text{S}_2\text{O}_3^{2-}$) is in turn putatively oxidized by the Sox multi-enzyme system [5,139,140], which has been described in another member of the Aquificaceae, *H. thermophilus* [141]. The Sox system has also been described in other thermophiles that share with *A. aeolicus* both the equivalent set of Sox genes in their genomes, and the characteristic of producing cytoplasmic sulfur globules under certain growth conditions [142]. Finally, rhodanese possesses several rhodanese complexes possibly involved in cyanide detoxification [143,144], as well as an ATP sulfurylase [145] possibly involved in sulfite oxidation [5]. None of these compounds has been shown to be able to replace O_2 as sole terminal electron acceptor, and due to the general complexity of sulfur chemistry, its roles in *A. aeolicus* energy metabolism remain to be fully mapped out [5].

Electrons as transferred from H_2 into core metabolism at three main points, Hydrogenases I, II & III, from where they enter the membrane quinone pool (Hydrogenase I, II) or are directly transferred to ferredoxin in the cytoplasm (soluble Hydrogenase III) [146]. In the Hydrogenase I respiratory chain, the quinones subsequently transfer the electrons to a cytochrome *bc*₁ complex, which then reduces O_2 to water [130,147,148]. In the Hydrogenase II respiratory chain, the quinones instead transfer the electrons to the sulfur reductase complex, which reduces elemental sulfur (and perhaps tetrathionate) and produces H_2S [133,138]. H_2S can then be re-oxidized by a sulfide quinone reductase complex that transfers the electrons through quinones (and a

cytochrome *bc*₁ complex) into oxygen, which is reduced to water [131,133,135].

Ultimately the energy metabolism of *A. aeolicus* produces a set of reductant carriers in the cytoplasm, which then drive its biosynthetic machinery. These reductants include nicotinamides (NADH, NADPH), flavins (FADH₂), and ferredoxins, while in special cases quinones can also directly driving metabolic conversions at the membrane surface (such as fumarate reduction to succinate). Finally, the bioenergetics of the electron transport chains generates a proton-motive force across the membrane, which *A. aeolicus* harvests by using an ATP synthase to generate ATP [149].

Outlook and Future Directions

***Aquifex aeolicus* as a model system for early metabolic evolution.** We have used phylometabolic analysis to reconstruct the whole-genome metabolic network of *A. aeolicus*, and have shown that it uses the likely ancestral pathways within many sub-systems. Perhaps even more important than whether this organism represents the “most ancestral” known metabolism or not, is that comparing the evolutionary trajectories of its sub-systems to those of other later branching bacteria often identifies the *direction* of evolutionary change going from the earliest cells to modern forms. This can help us infer properties of those earliest cells, even if the evolutionary change has already proceeded in *A. aeolicus*.

For example, our analyses have shown that *A. aeolicus* uses variants of both the rTCA cycle and the branched chain amino acid biosynthesis pathways reconstructed to be more ancestral than those used by other organism, which make greater use of parallel sequences of homologous local functional group chemistry. Combined with a general observation that selection appears to optimize the kinetics of enzymes by increasing their substrate specificity, this allows us to hypothesize that ancestral life forms could have regulated those sub-networks with more promiscuous enzymes and a smaller genome, with evolution then producing the forms used by *A. aeolicus*.

There are indications that evolution may have modified some *A. aeolicus* metabolic sub-networks to such a degree that they are not closer to the ancestral form of those sub-network than other extant forms. Particularly notable among these is the apparent obligatory use of oxygen as the terminal electron acceptor, despite the fact that it is generally accepted that oxygen did not rise significantly in Earth’s atmosphere until around 2.5 GYa [150,151]. Rather than using this single observation to place the Aquificales as a late-branching clade that arose during or after the rise of oxygen, we instead argue that all evidence combined suggests that the use of oxygen is a later adaptation within energy metabolism that emerged in a sub-set of Aquificales.

Several observations support this interpretation. Many Aquificales do not use oxygen, instead using a fully anoxic metabolism relying on nitrate, elemental sulfur or thiosulfate as their terminal electron acceptors, while yet others can switch between anaerobic and microaerophile lifestyles [16,136]. *Aquifex pyrophilus*, the closest known relative to *A. aeolicus*, can use both nitrate and oxygen as its terminal electron acceptor [152], which as discussed before remains a possibility in *A. aeolicus* as well. As also discussed, *A. aeolicus* can use various sulfur compounds as intermediate electron acceptors, but not as the terminal acceptor, possibly because of the tight coupling between the components of its respiratory super-complexes. Finally, the rTCA cycle has several oxygen-sensitive components, and all organisms using this carbon-fixation pathway (including all members of the Aquificales, Chlorobiales, ϵ -proteobacteria and Nitrospirae) are anaerobes or microaerophiles [53].

Altogether these observations lead to a plausible evolutionary scenario of Aquificale energy metabolism, from a fully anaerobic ancestral state toward diverse forms that cover the anaerobe-microaerophile spectrum. In this scenario, *A. aeolicus* could have sequentially added sub-components, with the oxygen handling parts added later, essentially encapsulating an otherwise anoxic chain within an oxygen-using chain. However, unlike most other Aquificales that also use oxygen, *A. aeolicus* seems to have lost the ability to use other intermediates within its electron transport chains as the terminal acceptor, possibly because of the tight integration of the respiratory chains (or simply because the right combination of experimental parameters have not yet been found). If this hypothesis is correct, then comparative genomic and experimental analyses of Aquificale energy metabolisms remains a key area of research for understanding the sequences of innovations that drove their evolutionary diversification.

There is likely no single organism today that best represents the earliest cells on earth, because all cells alive today are the product of a several billion year evolutionary process. However, the discussion about oxygen use in *A. aeolicus* was partly meant to highlight that different sub-components and sub-networks within bacterial (or archaeal) lineages may evolve independently from the rest of the system in response to specific selective pressures in the environment. Thus, some organisms may give more insights into the earliest forms of life, because their environments have changed less over the course of Earth history, and because they retain a larger number of sub-components that resemble reconstructed ancestral forms of those sub-components. In this regard we suggest that *A. aeolicus* is as good as any, and better than most, as a model system for studying the earliest metabolic evolution of life on Earth.

General principles of metabolic evolution. By reconstructing the evolutionary sequences among pathways in *A. aeolicus*, we identified several general driving forces of metabolic evolution. We have highlighted throughout the way selection for improved kinetics and/or improved thermodynamics can shape a metabolic network. Comparing different sub-networks further reveals trade-offs in the costs versus the benefits of these innovations, which appears to depend strongly on the relative mass flux density of sub-networks. Extending these analyses to other metabolic sub-networks, and to the evolutionary history of other organisms will improve our understanding of how tradeoffs between performance gains and their associated costs generally contributed to fitness in the earliest stages of cellular life.

Supporting Information

Text S1 Additional details on the *A. aeolicus* metabolic network reconstruction and curation.
(PDF)

File S1 The whole-organism metabolic network of *A. aeolicus* in SBML format [36].
(XML)

Acknowledgments

We gratefully acknowledge two anonymous reviewers for helpful feedback on our manuscript.

Author Contributions

Conceived and designed the experiments: RB ES. Performed the experiments: RB. Analyzed the data: RB. Wrote the paper: RB ES.

References

- Braakman R, Smith E (2013) The compositional and evolutionary logic of metabolism. *Physical Biology* 10: 011001.
- Braakman R, Smith E (2012) The emergence and early evolution of biological carbon fixation. *PLoS Computational Biology* 8: e1002455.
- Huber R, Eder W (2006) Aquificales. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E, editors, *The Prokaryotes*, Springer New York. 925–938.
- Reysenbach AL, Shock E (2002) Merging genomes with geochemistry in hydrothermal ecosystems. *Science* 296: 1077–1082.
- Guiral M, Prunetti L, Aussignargues C, Ciaccavava A, Infossi P, et al. (2012) The hyperthermophilic bacterium *Aquifex aeolicus*: From respiratory pathways to extremely resistant enzymes and biotechnological applications. *Advances in Microbial Physiology* 61: 125.
- Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. *EMBO Journal* 5: 823–826.
- Schnoes AM, Brown SD, Dodevski I, Babbitt PC (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS computational biology* 5: e1000605.
- Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, et al. (2013) A large-scale evaluation of computational protein function prediction. *Nature Methods* 10: 221–227.
- Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology* 7: 129–143.
- Henry CS, Zimmer JF, Cohoon MP, Stevens RL (2009) iBsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. *Genome Biology* 10: R69.
- Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, et al. (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 392: 353–358.
- Burggraf S, Olsen GJ, Stetter KO, Woese CR (1992) A phylogenetic analysis of *Aquifex pyrophilus*. *Systematic and Applied Microbiology* 15: 352–356.
- Pace NR (1997) A molecular view of microbial diversity and the biosphere. *Science* 276: 734–740.
- Griffiths E, Gupta RS (2004) Signature sequences in diverse proteins provide evidence for the late divergence of the Order Aquificales. *International Microbiology* 7: 41–52.
- Boussau B, Guéguen L, Gouy M (2008) Accounting for horizontal gene transfers explains conflicting hypotheses regarding the position of Aquificales in the phylogeny of bacteria. *BMC Evolutionary Biology* 8: 272: 1–18.
- Sievert S, Vetriani C (2012) Chemoautotrophy at deep-sea vents: Past, present, and future. *Oceanography* 25: 218–233.
- Polz MF, Alm EJ, Hanage WP (2013) Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in Genetics* 29: 170–175.
- Zhaxybayeva O, Swithers KS, Lapierre P, Fournier GP, Bickhart DM, et al. (2009) On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proceedings of the National Academy of Sciences USA* 106: 5865–5870.
- Jensen RA (1976) Enzyme recruitment in evolution of new function. *Annual Review of Microbiology* 30: 409–425.
- O'Brien PJ, Herschlag D (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chemistry & Biology* 6: R91–R105.
- Copley SD (2003) Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Current Opinions in Chemical Biology* 7: 265–272.
- Khersonsky O, Tawfik DS (2010) Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annual Review of Biochemistry* 79: 471–505.
- Tawfik DS, et al. (2010) Messy biology and the origins of evolutionary innovations. *Nature Chemical Biology* 6: 692.
- Kim J, Kershner JP, Novikov Y, Shoemaker RK, Copley SD (2010) Three serendipitous pathways in *E. coli* can bypass a block in pyridoxal-50-phosphate synthesis. *Molecular Systems Biology* 6: 436: 1–13.
- Khersonsky O, Malitsky S, Rogachev I, Tawfik DS (2011) Role of chemistry versus substrate binding in recruiting promiscuous enzyme functions. *Biochemistry* 50: 2683–2690.
- Nam H, Lewis NE, Lerman JA, Lee DH, Chang RL, et al. (2012) Network context and selection in the evolution to enzyme specificity. *Science* 337: 1101–1104.
- Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, et al. (2011) The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50: 4402–4410.
- Fondi M, Brilli M, Emiliani G, Paffetti D, Fani R (2007) The primordial metabolism: an ancestral interconnection between leucine, arginine, and lysine biosynthesis. *BMC Evolutionary Biology* (Suppl 2): S3.
- Peretó J (2012) Out of fuzzy chemistry: from prebiotic chemistry to metabolic networks. *Chemical Society Reviews* 41: 5394–5403.
- Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, et al. (1999) Detecting protein function and protein-protein interactions from genome sequences. *Science* 285: 751–753.
- Snel B, Bork P, Huynen M, et al. (2000) Genome evolution-gene fusion versus gene fission. *Trends in Genetics* 16: 9–11.
- Kummerfeld SK, Teichmann SA, et al. (2005) Relative rates of gene fusion and fission in multidomain proteins. *Trends in Genetics* 21: 25.
- Pfeiffer T, Schuster S, Bonhoeffer S (2001) Cooperation and competition in the evolution of ATP-producing pathways. *Science* 292: 504–507.
- Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, et al. (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology* 28: 977–982.
- The UniProt Consortium (2011) Ongoing and future developments at the universal protein resource. *Nucleic Acids Research* 39: D214–D219.
- Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19: 524–531.
- Zhang Y, Thiele I, Weekes D, Li Z, Jaroszewski L, et al. (2009) Three-dimensional structural view of the central metabolic network of *Thermotoga maritima*. *Science* 325: 1544–1549.
- White SW, Zheng J, Zhang YM, Rock CO (2005) The structural biology of type II fatty acid biosynthesis. *Annual Review of Biochemistry* 74: 791–831.
- Carballeira NM, Reyes M, Sostre A, Huang H, Verhagen M, et al. (1997) Unusual fatty acid compositions of the hyperthermophilic archaeon *Pyrococcus furiosus* and the bacterium *Thermotoga maritima*. *Journal of Bacteriology* 179: 2766–2768.
- Beh M, Strauss G, Huber R, Stetter KO, Fuchs G (1993) Enzymes of the reductive citric acid cycle in the autotrophic eubacterium *Aquifex pyrophilus* and in the archaeobacterium *Thermoproteus neutrophilus*. *Archives of Microbiology* 160: 306–311.
- Hügler M, Huber H, Molyneux SJ, Vetriani C, Seibert SM (2007) Autotrophic CO₂ fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum Aquificae: evidence for two ways of citrate cleavage. *Environmental Microbiology* 9: 81–92.
- Smith E, Morowitz HJ (2004) Universality in intermediary metabolism. *Proceeding of the National Academy of Sciences USA* 101: 13168–13173.
- Maden BEH (2000) Tetrahydrofolate and tetrahydromethanopterin compared: functionally distinct carriers in C1 metabolism. *Biochemical Journal* 350: 609–629.
- Ragsdale SW, Clark JE, Ljungdahl LG, Lundie LL, Drake HL (1983) Properties of purified carbon monoxide dehydrogenase from *Clostridium thermoacetatum*, a nickel, iron-sulfur protein. *Journal of Biological Chemistry* 258: 2364–2369.
- Stover P, Schirch V (1993) The metabolic role of leucovorin. *Trends in Biochemical Sciences* 18: 102–106.
- Huang T, Schirch V (1995) Mechanism for the coupling of ATP hydrolysis to the conversion of 5-formyltetrahydrofolate to 5,10-methylenetetrahydrofolate. *Journal of Biological Chemistry* 270: 22296–22300.
- von Wettstein D, Gough S, Kannangara CG (1995) Chlorophyll biosynthesis. *The Plant Cell* 7: 1039–1057.
- Benner SA, Ellington AD, Tauer A (1989) Modern metabolism as a palimpsest of the RNA world. *Proceeding of the National Academy of Sciences USA* 18: 7054–7058.
- Aoshima M, Ishii M, Igarashi Y (2004) A novel enzyme, citryl-CoA synthetase, catalysing the first step of the citrate cleavage reaction in *Hydrogenobacter thermophilus* TK-6. *Molecular Microbiology* 52: 751–761.
- Aoshima M, Ishii M, Igarashi Y (2004) A novel enzyme, citryl-CoA lyase, catalysing the second step of the citrate cleavage reaction in *Hydrogenobacter thermophilus* TK-6. *Molecular Microbiology* 52: 763–770.
- Aoshima M, Ishii M, Igarashi Y (2004) A novel biotin protein required for reductive carboxylation of 2-oxoglutarate by isocitrate dehydrogenase in *Hydrogenobacter thermophilus* TK-6. *Molecular Microbiology* 51: 791–798.
- Aoshima M, Igarashi Y (2006) A novel oxalosuccinate-forming enzyme involved in the reductive carboxylation of 2-oxoglutarate in *Hydrogenobacter thermophilus* TK-6. *Molecular Microbiology* 62: 748–759.
- Hügler M, Seibert SM (2011) Beyond the calvin cycle: Autotrophic carbon fixation in the ocean. *Annual Review of Marine Science* 3: 261–289.
- Aoshima M, Igarashi Y (2008) Nondecarboxylating and decarboxylating isocitrate dehydrogenases: oxalosuccinate reductase as an ancestral form of isocitrate dehydrogenase. *Journal of Bacteriology* 190: 2050–2055.
- Bar-Even A, Flamholz A, Noor E, Milo R (2012) Thermodynamic constraints shape the structure of carbon fixation pathways. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1817: 1646–1659.
- Miller SL, Smith-Magowan D (1990) The thermodynamics of the Krebs cycle and related compounds. *Journal of Physical and Chemical Reference Data* 19: 1049–1073.
- Wolfenden R, Lewis Jr CA, Yuan Y (2011) Kinetic challenges facing oxalate, malonate, acetoacetate and oxaloacetate decarboxylases. *Journal of the American Chemical Society* 133: 5683–5685.
- Buchanan BB, Arnold DI (1990) A reverse Krebs cycle in photosynthesis: Consensus at last. *Photosynthesis Research* 24: 47–53.
- Lücker S, Wagner M, Maixner F, Pelletier E, Koch H, et al. (2010) A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proceeding of the National Academy of Sciences USA* 107: 13479–13484.

60. Hügl M, Wirsén CO, Fuchs G, Taylor CD, Sievert SM (2005) Evidence for autotrophic CO₂ fixation via the reductive tricarboxylic acid cycle by members of the ϵ subdivision of Proteobacteria. *Journal of Bacteriology* 187: 3020–3027.
61. Fuchs G (2011) Alternative pathways of carbon dioxide fixation: Insights into the early evolution of life? *Annual Review of Microbiology* 65: 631–658.
62. Scapin G, Blanchard J (1998) Enzymology of bacterial lysine biosynthesis. *Advances in enzymology and related areas of molecular biology* 72: 279.
63. Bentley R, Haslam E (1990) The shikimate pathway a metabolic tree with many branches. *Critical Reviews in Biochemistry and Molecular Biology* 25: 307–384.
64. Umbarger HE, Brown B (1957) Threonine deamination in *Escherichia coli* II: Evidence for two l-threonine deaminases. *Journal of Bacteriology* 73: 105.
65. Charon NW, Johnson RC, Peterson D (1974) Amino acid biosynthesis in the spirochete *Leptospira*: evidence for a novel pathway of isoleucine biosynthesis. *Journal of Bacteriology* 117: 203–211.
66. Xu H, Zhang Y, Guo X, Ren S, Staempi AA, et al. (2004) Isoleucine biosynthesis in *Leptospira interrogans* serotype lai strain 56601 proceeds via a threonine-independent pathway. *Journal of Bacteriology* 186: 5400–5409.
67. Eikmanns B, Linder D, Thauer RK (1983) Unusual pathway of isoleucine biosynthesis in *Methanobacterium thermoautotrophicum*. *Archives of Microbiology* 136: 111–113.
68. Hochuli M, Patzelt H, Oesterhelt D, Wüthrich K, Szyperski T (1999) Amino acid biosynthesis in the halophilic archaeon *Haloarcula hispanica*. *Journal of Bacteriology* 181: 3226–3237.
69. Howell DM, Xu H, White RH (1999) (r)-citramalate synthase in methanogenic archaea. *Journal of Bacteriology* 181: 331–333.
70. Drevland RM, Waheed A, Graham DE (2007) Enzymology and evolution of the pyruvate pathway to 2-oxobutyrate in *Methanocaldococcus jannaschii*. *Journal of Bacteriology* 189: 4391–4400.
71. Schäfer S, Barkowski C, Fuchs G (1986) Carbon assimilation by the autotrophic thermophilic archaeobacterium *Thermoproteus neutrophilus*. *Archives of Microbiology* 146: 301–308.
72. Jahn U, Huber H, Eisenreich W, Hügl M, Fuchs G (2007) Insights into the autotrophic CO₂ fixation pathway of the archaeon *Ignicoccus hospitalis*: comprehensive analysis of the central carbon metabolism. *Journal of Bacteriology* 189: 4108–4119.
73. Feng X, Mouttaki H, Lin L, Huang R, Wu B, et al. (2009) Characterization of the central metabolic pathways in *Thermoanaerobacter* sp. strain x514 via isotopomer-assisted metabolite analysis. *Applied and Environmental Microbiology* 75: 5001–5008.
74. Tang KH, Feng X, Zhuang WQ, Alvarez-Cohen L, Blankenship RE, et al. (2010) Carbon flow of heliobacteria is related more to *Clostridia* than to the green sulfur bacteria. *Journal of Biological Chemistry* 285: 35104–35112.
75. Tang YJ, Yi S, Zhuang WQ, Zinder SH, Keasling JD, et al. (2009) Investigation of carbon metabolism in *Dehalococcoides ethenogenes* strain 195 by use of isotopomer and transcriptomic analyses. *Journal of Bacteriology* 191: 5224–5231.
76. Feng X, Tang KH, Blankenship RE, Tang YJ (2010) Metabolic flux analysis of the mixotrophic metabolisms in the green sulfur bacterium *Chlorobaculum tepidum*. *Journal of Biological Chemistry* 285: 39544–39550.
77. Wu B, Zhang B, Feng X, Rubens JR, Huang R, et al. (2010) Alternative isoleucine synthesis pathway in cyanobacterial species. *Microbiology* 156: 596–602.
78. Risso C, Van Dien SJ, Orloff A, Lovley DR, Coppi MV (2008) Elucidation of an alternate isoleucine biosynthesis pathway in *Geobacter sulfurreducens*. *Journal of Bacteriology* 190: 2266–2274.
79. Tang KH, Feng X, Tang YJ, Blankenship RE (2009) Carbohydrate metabolism and carbon fixation in *Roseobacter denitrificans* och114. *PLoS One* 4: e7233.
80. McKinlay JB, Harwood CS (2010) Carbon dioxide fixation as a central redox cofactor recycling mechanism in bacteria. *Proceedings of the National Academy of Sciences USA* 107: 11669–11675.
81. Xie G, Forst C, Bonner C, Jensen RA, et al. (2002) Significance of two distinct types of tryptophan synthase beta chain in bacteria, archaea and higher plants. *Genome Biology* 3.
82. Wiegel J, Tanner R, Rainey FA, et al. (2006) An introduction to the family *Clostridiaceae*. *Prokaryotes* 4: 654–678.
83. Urbach E, Scanlan DJ, Distel DL, Waterbury JB, Chisholm SW (1998) Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (cyanobacteria). *Journal of Molecular Evolution* 46: 188–201.
84. Palenik B, Brahmasha B, Larimer F, Land M, Hauser L, et al. (2003) The genome of a motile marine *Synechococcus*. *Nature* 424: 1037–1042.
85. Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, et al. (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424: 1042–1047.
86. Omelchenko M, Wolf Y, Gaidamakova E, Matrosova V, Vasilenko A, et al. (2005) Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes of adaptation to thermophily and radiation resistance. *BMC Evolutionary Biology* 5: 57.
87. Umbarger H (1978) Amino acid biosynthesis and its regulation. *Annual Review of Biochemistry* 47: 533–606.
88. Grimminger H, Umbarger H (1979) Acetohydroxy acid synthase I of *Escherichia coli*: purification and properties. *Journal of Bacteriology* 137: 846–853.
89. Barak Z, Chipman DM, Gollop N (1987) Physiological implications of the specificity of acetohydroxy acid synthase isozymes of enteric bacteria. *Journal of Bacteriology* 169: 3750–3756.
90. Srinivasan V, Morowitz HJ (2009) The canonical network of autotrophic intermediary metabolism: Minimal metabolome of a reductive chemoautotroph. *Biological Bulletin* 216: 126–130.
91. Kikuchi G (1973) The glycine cleavage system: Composition, reaction mechanism, and physiological significance. *Molecular and Cellular Biochemistry* 1: 169–187.
92. Cronan JE, Zhao X, Jiang Y (2005) Function, attachment and synthesis of lipoic acid in *Escherichia coli*. *Advances in Microbial Physiology* 50: 103–146.
93. Zhao X, Miller JR, Jiang Y, Marletta MA, Cronan JE (2003) Assembly of the covalent linkage between lipoic acid and its cognate enzymes. *Chemistry & Biology* 10: 1293–1302.
94. Christensen QH, Cronan JE (2010) Lipoic acid synthesis: A new family of octanoyltransferases generally annotated as lipoate protein ligases. *Biochemistry* 49: 10024–10036.
95. Martin N, Christensen QH, Mansilla MC, Cronan JE, de Mendoza D (2011) A novel two-gene requirement for the octanoyltransfer reaction of *Bacillus subtilis* lipoic acid biosynthesis. *Molecular Microbiology* 80: 335–349.
96. Christensen QH, Martin N, Mansilla MC, de Mendoza D, Cronan JE (2011) A novel amidotransferase required for lipoic acid cofactor assembly in *Bacillus subtilis*. *Molecular Microbiology* 80: 350–363.
97. Booker SJ (2004) Unraveling the pathway of lipoic acid biosynthesis. *Chemistry & Biology* 11: 10–12.
98. Hermes FA, Cronan JE (2009) Scavenging of cytosolic octanoic acid by mutant LplA lipoate ligases allows growth of *Escherichia coli* strains lacking the LipB octanoyltransferase of lipoic acid synthesis. *Journal of Bacteriology* 191: 6796–6803.
99. Rock CO (2009) Opening a new path to lipoic acid. *Journal of Bacteriology* 191: 6782–6784.
100. Christensen QH, Cronan JE (2009) The *Thermoplasma acidophilum* LplA-LplB complex defines a new class of bipartite lipoate-protein ligases. *Journal of Biological Chemistry* 284: 21317–21326.
101. Cicchillo RM, Iwig DF, Jones AD, Nesbitt NM, Baleanu-Gogonea C, et al. (2004) Lipoyl synthase requires two equivalents of S-adenosyl-L-methionine to synthesize one equivalent of lipoic acid. *Biochemistry* 43: 6378–6386.
102. Fitzpatrick T, Amrhein N, Kappes B, Macheroux P, Tewes I, et al. (2007) Two independent routes of de novo vitamin B6 biosynthesis: not that different after all. *Biochemical Journal* 407: 1–13.
103. Austin SM, Waddell TG (1999) Prebiotic synthesis of vitamin B6-type compounds. *Origins of Life and Evolution of the Biosphere* 29: 287–296.
104. Morowitz HJ, Srinivasan V, Smith E (2006) The swiss army knife of biological catalysis: A compact toolkit of organic functional groups. *Complexity* 11: 9–10.
105. Burns KE, Xiang Y, Kinsland CL, McLafferty FW, Begley TP (2005) Reconstitution and biochemical characterization of a new pyridoxal-5'-phosphate biosynthetic pathway. *Journal of the American Chemical Society* 127: 3682–3683.
106. Raschle T, Amrhein N, Fitzpatrick TB (2005) On the two components of pyridoxal 5'-phosphate synthase from *Bacillus subtilis*. *Journal of Biological Chemistry* 280: 32291–32300.
107. Ehrenshaft M, Bilski P, Li MY, Chignell CF, Daub ME (1999) A highly conserved sequence is a novel gene involved in de novo vitamin B6 biosynthesis. *Proceedings of the National Academy of Sciences USA* 96: 9374–9378.
108. Mittenhuber G, et al. (2001) Phylogenetic analyses and comparative genomics of vitamin B6 (pyridoxine) and pyridoxal phosphate biosynthesis pathways. *Journal of Molecular Microbiology and Biotechnology* 3: 1–20.
109. Boucher Y, Doolittle WF (2000) The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Molecular microbiology* 37: 703–716.
110. Infossi P, Lojou E, Chauvin JP, Herbette G, Brugna M, et al. (2010) *Aquifex aeolicus* membrane hydrogenase for hydrogen biooxidation: Role of lipids and physiological partners in enzyme stability and activity. *International Journal of Hydrogen Energy* 35: 10778–10789.
111. Ishii M, Kawasumi T, Igarashi Y, Kodama T, Minoda Y (1987) 2-methylthio-1, 4-naphthoquinone, a unique sulfur-containing quinone from a thermophilic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus*. *Journal of Bacteriology* 169: 2380–2384.
112. Shima S, Suzuki KI (1993) *Hydrogenobacter acidophilus* sp. nov., a thermoacidophilic, aerobic, hydrogen-oxidizing bacterium requiring elemental sulfur for growth. *International Journal of Systematic Bacteriology* 43: 703–708.
113. Stohr R, Waberski A, Völker H, Tindall BJ, Thomm M (2001) *Hydrogenothermus marinus* gen. nov., sp. nov., a novel thermophilic hydrogen-oxidizing bacterium, recognition of *Calderobacterium hydrogenophilum* as a member of the genus *Hydrogenobacter* and proposal of the reclassification of *Hydrogenobacter acidophilus* as *Hydrogenobaculum acidophilum* gen. nov., comb. nov., in the phylum *Hydrogenobacter/Aquifex*. *International journal of systematic and evolutionary microbiology* 51: 1853–1862.
114. Nitschke W, Kramer D, Riedel A, Liebl U (1995) From naphtho- to benzoquinones - (r)evolutionary reorganizations of electron transfer chains. In: Mathis P, editor, *Photosynthesis: from Light to the Biosphere*, vol. 1, Dordrecht: Kluwer Academic Press. 945–950.

115. Schütz M, Brugna M, Lebrun E, Baymann F, Huber R, et al. (2000) Early evolution of cytochrome bc complexes. *Journal of Molecular Biology* 300: 663–675.
116. Schoepp-Cothenet B, Lieutaud C, Baymann F, Verméglio A, Friedrich T, et al. (2005) Menaquinone as a pool quinone in a purple bacterium. *Proceeding of the National Academy of Sciences USA* 106: 8549–8554.
117. Iverson TM, Luna-Chavez C, Cecchini G, Rees DC (1999) Structure of the *Escherichia coli* fumarate reductase respiratory complex. *Science* 284: 1961–1966.
118. Zhang Y, Morar M, Ealick SE (2008) Structural biology of the purine biosynthetic pathway. *Cellular and Molecular Life Sciences* 65: 3699–3724.
119. Heppel LA, Hilmoe RJ (1951) Purification and properties of 5-nucleotidase. *Journal of Biological Chemistry* 188: 665–676.
120. Berg P, Joklik WK (1954) Enzymatic phosphorylation of nucleoside diphosphates. *Journal of Biological Chemistry* 210: 657–672.
121. White RH (1997) Purine biosynthesis in the domain archaea without folates or modified folates. *Journal of Bacteriology* 179: 3374–3377.
122. Marolewski A, Smith JM, Benkovic SJ (1994) Cloning and characterization of a new purine biosynthetic enzyme: a non-folate glycinamide ribonucleotide transformylase from *E. coli*. *Biochemistry* 33: 2531–2537.
123. Ahuja A, Purcarea C, Guy HI, Evans DR (2001) A novel carbamoyl-phosphate synthetase from *Aquifex aeolicus*. *Journal of Biological Chemistry* 276: 45694–45703.
124. Ahuja A, Purcarea C, Ebert R, Sadecki S, Guy HI, et al. (2004) *Aquifex aeolicus* dihydroorotase association with aspartate transcarbamoylase switches on catalytic activity. *Journal of Biological Chemistry* 279: 53136–53144.
125. Schleifer KH, Kandler O (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological Reviews* 36: 407.
126. Jahnke LL, Eder W, Huber R, Hope JM, Hinrichs KU, et al. (2001) Signature lipids and stable carbon isotope analyses of octopus spring hyperthermophilic communities compared with those of aquificales representatives. *Applied and Environmental Microbiology* 67: 5179–5189.
127. Lengeler JW, Drews G, Schlegel HG (1999) *Biology of the Prokaryotes*. New York: Blackwell Science.
128. Zhang YM, Rock CO (2008) Membrane lipid homeostasis in bacteria. *Nature Reviews Microbiology* 6: 222–233.
129. Valentine DL (2007) Adaptations to energy stress dictate the ecology and evolution of the archaea. *Nature Reviews Microbiology* 5: 316–323.
130. Peng G, Fritsch G, Zickermann V, Schaagger H, Mentele R, et al. (2003) Isolation, characterization and electron microscopic single particle analysis of the NADH : ubiquinone oxidoreductase (complex I) from the hyperthermophilic eubacterium *Aquifex aeolicus*. *Biochemistry* 42: 3032–3039.
131. Guiral M, Prunetti L, Lignon S, Lebrun R, Moinier D, et al. (2009) New insights into the respiratory chains of the chemolithoautotrophic and hyperthermophilic bacterium *Aquifex aeolicus*. *Journal of Proteome Research* 8: 1717–1730.
132. Trouillard M, Meunier B, Rappaport F (2011) Questioning the functional relevance of mitochondrial supercomplexes by time-resolved analysis of the respiratory chain. *Proceedings of the National Academy of Sciences USA* 108: E1027–E1034.
133. Nubel T, Klughammer C, Huber R, Hauska G, Schutz M (2000) Sulfide : quinone oxidoreductase in membranes of the hyperthermophilic bacterium *Aquifex aeolicus* (VF5). *Archives of Microbiology* 173: 233–244.
134. Pelletier N, Leroy G, Guiral M, Giudici-Orticoni MT, Aubert C (2008) First characterisation of the active oligomer form of sulfur oxygenase reductase from the bacterium *Aquifex aeolicus*. *Extremophiles* 12: 205–215.
135. Prunetti L, Infossi P, Brugna M, Ebel C, Giudici-Orticoni MT, et al. (2010) New functional sulfide oxidase-oxygen reductase supercomplex in the membrane of the hyperthermophilic bacterium *Aquifex aeolicus*. *Journal of Biological Chemistry* 285: 41815–41826.
136. Vetriani C, Speck MD, Ellor SV, Lutz RA, Starovoytov V (2004) *Thermovibrio ammonificans* sp. nov., a thermophilic, chemolithotrophic, nitrate-ammonifying bacterium from deep-sea hydrothermal vents. *International Journal of Systematic and Evolutionary Microbiology* 54: 175–181.
137. Wald G (1962) Life in the second and third periods: or why phosphorus and sulfur for high-energy bonds. In: Kasha M, Pullman B, editors, *Horizons in biochemistry*. New York: Academic Press, 127–142.
138. Guiral M, Tron P, Aubert C, Gloter A, Iobbi-Nivol C, et al. (2005) A membrane-bound multienzyme, hydrogen-oxidizing, and sulfur-reducing complex from the hyperthermophilic bacterium *Aquifex aeolicus*. *Journal of Biological Chemistry* 280: 42004–42015.
139. Verte F, Kostanjevecki V, De Smet L, Meyer T, Cusanovich M, et al. (2002) Identification of a thiosulfate utilization gene cluster from the green phototrophic bacterium *Chlorobium limicola*. *Biochemistry* 41: 2932–2945.
140. Ghosh W, Mallick S, DasGupta SK (2009) Origin of the sox multienzyme complex system in ancient thermophilic bacteria and coevolution of its constituent proteins. *Research in Microbiology* 160: 409–420.
141. Sano R, Kameya M, Wakai S, Arai H, Igarashi Y, et al. (2010) Thiosulfate oxidation by a thermoneutrophilic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus*. *Bioscience, Biotechnology, and Biochemistry* 74: 892–894.
142. Miyake D, Ichiki Si, Tanabe M, Oda T, Kuroda H, et al. (2007) Thiosulfate oxidation by a moderately thermophilic hydrogen-oxidizing bacterium, *Hydrogenophilus thermoluteolus*. *Archives of Microbiology* 188: 199–204.
143. Giuliani MC, Tron P, Leroy G, Aubert C, Tauc P, et al. (2007) A new sulfurtransferase from the hyperthermophilic bacterium *Aquifex aeolicus*. *FEBS Journal* 274: 4572–4587.
144. Giuliani MC, Jourlin-Castelli C, Leroy G, Hachani A, Giudici-Orticoni MT (2010) Characterization of a new periplasmic single-domain rhodanese encoded by a sulfur-regulated gene in a hyperthermophilic bacterium *Aquifex aeolicus*. *Biochimie* 92: 388–397.
145. Yu Z, Lansdon EB, Segel IH, Fisher AJ (2007) Crystal structure of the bifunctional ATP sulfurylase: APS kinase from the chemolithotrophic thermophile *Aquifex aeolicus*. *Journal of Molecular Biology* 365: 732–743.
146. Guiral M, Aubert T, Giudici-Orticoni M (2005) Hydrogen metabolism in the hyperthermophilic bacterium *Aquifex aeolicus*. *Biochemical Society Transactions* 33: 22–24.
147. Brugna-Guiral M, Tron P, Nitschke W, Stetter KO, Burlat B, et al. (2003) [NiFe] hydrogenases from the hyperthermophilic bacterium *Aquifex aeolicus*: properties, function, and phylogenetics. *Extremophiles* 7: 145–157.
148. Schutz M, Schoepp-Cothenet B, Lojou E, Woodstra M, Lexa D, et al. (2003) The naphthoquinol oxidizing cytochrome bc1 complex of the hyperthermophilic knallgasbacterium *Aquifex aeolicus*: Properties and phylogenetic relationships. *Biochemistry* 42: 10800–10808.
149. Peng G, Bostina M, Radermacher M, Rais I, Karas M, et al. (2006) Biochemical and electron microscopic characterization of the F1F0 ATP synthase from the hyperthermophilic eubacterium *Aquifex aeolicus*. *FEBS letters* 580: 5934–5940.
150. Buick R (2008) When did oxygenic photosynthesis evolve? *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 2731–2743.
151. Kump LR (2008) The rise of atmospheric oxygen. *Nature* 451: 277–278.
152. Huber R, Wilharm T, Huber D, Trincone A, Burggraf S, et al. (1992) *Aquifex pyrophilus* gen. nov. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. *Systematic and Applied Microbiology* 15: 340–351.