

Phylogenomic Analysis and Predicted Physiological Role of the Proton-Translocating NADH:Quinone Oxidoreductase (Complex I) Across Bacteria

Melanie A. Spero,^{a,b} Frank O. Aylward,^{a,b} Cameron R. Currie,^{a,b}  Timothy J. Donohue^{a,b}

Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin, USA^a; Great Lakes Bioenergy Research Center, Madison, Wisconsin, USA^b

ABSTRACT The proton-translocating NADH:quinone oxidoreductase (complex I) is a multisubunit integral membrane enzyme found in the respiratory chains of both bacteria and eukaryotic organelles. Although much research has focused on the enzyme's central role in the mitochondrial respiratory chain, comparatively little is known about its role in the diverse energetic lifestyles of different bacteria. Here, we used a phylogenomic approach to better understand the distribution of complex I across bacteria, the evolution of this enzyme, and its potential roles in shaping the physiology of different bacterial groups. By surveying 970 representative bacterial genomes, we predict complex I to be present in ~50% of bacteria. While this includes bacteria with a wide range of energetic schemes, the presence of complex I is associated with specific lifestyles, including aerobic respiration and specific types of phototrophy (bacteria with only a type II reaction center). A phylogeny of bacterial complex I revealed five main clades of enzymes whose evolution is largely congruent with the evolution of the bacterial groups that encode complex I. A notable exception includes the gammaproteobacteria, whose members encode one of two distantly related complex I enzymes predicted to participate in different types of respiratory chains (aerobic versus anaerobic). Comparative genomic analyses suggest a broad role for complex I in reoxidizing NADH produced from various catabolic reactions, including the tricarboxylic acid (TCA) cycle and fatty acid beta-oxidation. Together, these findings suggest diverse roles for complex I across bacteria and highlight the importance of this enzyme in shaping diverse physiologies across the bacterial domain.

IMPORTANCE Living systems use conserved energy currencies, including a proton motive force (PMF), NADH, and ATP. The respiratory chain enzyme, complex I, connects these energy currencies by using NADH produced during nutrient breakdown to generate a PMF, which is subsequently used for ATP synthesis. Our goal is to better understand the role of complex I in bacteria, whose energetic diversity allows us to view its function in a range of biological contexts. We analyzed sequenced bacterial genomes to predict the presence, evolution, and function of complex I in bacteria. We identified five main classes of bacterial complex I and predict that different classes participate in different types of respiratory chains (aerobic and anaerobic). We also predict that complex I helps maintain a cellular redox state by reoxidizing NADH produced from central metabolism. Our findings suggest diverse roles for complex I in bacterial physiology, highlighting the need for future laboratory-based studies.

Received 19 March 2015 Accepted 20 March 2015 Published 14 April 2015

Citation Spero MA, Aylward FO, Currie CR, Donohue TJ. 2015. Phylogenomic analysis and predicted physiological role of the proton-translocating NADH:quinone oxidoreductase (complex I) across bacteria. *mBio* 6(2):00389-15. doi:10.1128/mBio.00389-15.

Editor Caroline S. Harwood, University of Washington

Copyright © 2015 Spero et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Timothy J. Donohue, tdonohue@bact.wisc.edu.

This article is a direct contribution from a Fellow of the American Academy of Microbiology.

Membrane-bound enzymes provide cells or organelles with the ability to acquire nutrients, remove toxic compounds, and perform metabolic functions crucial for growth and survival. Among such integral membrane enzymes are those within the respiratory and photosynthetic electron transport chains, which provide a vital means of connecting catabolism to energy conservation and other essential metabolic processes. This is exemplified by the first enzyme of the canonical aerobic respiratory chain, the proton-translocating NADH:quinone oxidoreductase (complex I).

Complex I catalyzes the reversible transfer of electrons from the soluble electron carrier NADH to membrane-bound quinone, coupling the energy of this reaction to the generation of a proton motive force (PMF) (1). This enzyme is central to energy conservation in most eukaryotes, where its action in mitochondria pro-

vides 40% of the PMF used for ATP synthesis (2). Thus, mutations in human complex I, which are the most common mitochondrial disorders, are associated with a range of pathological conditions and can be fatal (3). Additionally, mitochondrial complex I is a major source of reactive oxygen species (4, 5), which are implicated in the aging process and a number of diseases (6–8). While the role of complex I is well studied within the context of the mitochondrial respiratory chain, less is known about its physiological roles outside eukaryotes. We are interested in the role of complex I in the bacterial domain, where the great energetic diversity of these organisms allows us to view this enzyme in a range of biological contexts that suggests its breadth of function.

Bacterial respiratory chains and energetic lifestyles are more diverse than their eukaryotic counterparts. Individual bacterial

species can couple a large number of electron donors with the use of oxygen or other terminal electron acceptors, while phototrophic bacteria have dedicated energetic pathways that conserve energy from light (9). Bacterial complex I is composed of 14 different protein subunits (NuoA to NuoN), which represent the core enzyme, containing the minimal number of protein subunits and all of the cofactors required for enzyme activity (1). Because of the relative simplicity of the bacterial enzyme (mitochondrial complex I is composed of ~45 unique protein subunits [10]), many significant studies of bacterial complex I have shed light on enzyme structure and mechanism (2, 11–14), while few studies have addressed the physiological role of the enzyme.

These studies, however, demonstrate that bacterial complex I contributes to a range of growth conditions and metabolic processes. For example, in the facultative anaerobe *Escherichia coli*, complex I is not required for growth by aerobic respiration, although it is required to support anaerobic fumarate respiration (15). In contrast, the purple nonsulfur bacterium *Rhodobacter capsulatus* requires complex I for phototrophic growth, where the enzyme catalyzes the reverse reaction (compared to its function in aerobic respiration), using the PMF to drive NADH synthesis from quinol. Here, the ability of complex I to interconvert redox energy and electrochemical energy serves to prevent the overreduction of the quinone pool and to provide cellular reducing equivalents for CO₂ fixation (16, 17). Moreover, in the epsilon-proteobacterium *Campylobacter jejuni*, a modified version of complex I accepts electrons from flavodoxin rather than NADH, allowing it to use a different electron donor (18). Modified complex I enzymes or partial complex I operons are not limited to the epsilon-proteobacteria (19–23) and indicate that complex I has evolved to accommodate substantially different energetic and metabolic needs in various bacterial groups.

Though complex I activity shapes the bioenergetics of the cell, the distribution and phylogenetic diversity of this enzyme across bacteria are unclear. In this study, we took a phylogenomic approach to accomplish three goals. First, we examined 970 representative bacterial genomes to catalogue the diversity of bacteria predicted to have complex I and identified phylogenetic groups in which this enzyme was common. Second, we constructed multi-locus phylogenies of both complex I (NuoA to NuoN) and conserved housekeeping genes (see Materials and Methods) to identify the different clades of bacterial complex I and to determine the extent to which horizontal gene transfer (HGT), gene loss events, and vertical inheritance have shaped the distribution of complex I in bacteria. Third, we performed comparative genomic gene enrichment analysis to identify conserved biochemical pathways in bacterial genomes that are predicted to encode complex I, thereby revealing trends in the physiology of bacteria that contain this enzyme. Our results provide a framework for understanding the evolutionary dynamics and physiological implications of this bioenergetic enzyme within the diverse physiological contexts of different groups of bacteria.

RESULTS AND DISCUSSION

Complex I is predicted to be widespread and associated with specific bacterial lifestyles. We analyzed the distribution of the 14-subunit proton-translocating NADH:quinone oxidoreductase (complex I) in 1,058 representative genomes (970 bacterial and 88 archaeal) and predicted the enzyme to be widespread in bacteria (predicted in 52% of analyzed genomes) and mostly absent in

archaea (predicted in only one genome, *Halorhabdus utahensis* DMS 12940). The genes encoding complex I (*nuoA* to *nuoN*) were colocalized in 86% of the bacterial genomes where the enzyme was found, indicating they may be part of a polycistronic operon (as is the case for *Escherichia coli* [24]).

Our results predict complex I to be widespread across the bacterial phylogeny (Fig. 1A) and present in the genomes of bacteria with a wide set of metabolic and energetic schemes, including obligate aerobes, facultative anaerobes, strict anaerobes, obligate lithotrophs, methylotrophs, methanotrophs, and anoxic phototrophs, suggesting that complex I contributes to survival or energy conservation in a variety of lifestyles. Analysis of the project metadata associated with each genome revealed a number of trends in the distribution of complex I, such as its association with specific respiratory or energetic lifestyles, including aerobic respiration and bacteria with only a type II photosynthetic reaction center (Fig. 1B; chi-square test, $P < 0.0005$). Moreover, the majority of Gram-negative bacterial genomes (63%) were found to encode complex I, while the enzyme was found in only 25% of Gram-positive bacterial genomes (almost all of which were *Actinobacteria* species). Lastly, complex I was also more likely to be found in larger (>4 Mbp) bacterial genomes (Fig. 1B).

A small fraction (~3%) of bacterial genomes contained two operons predicted to encode distinct complex I isozymes (Table 1). These bacteria were mostly members of *Proteobacteria* but also included a member each of *Bacteroidetes* and *Gemmatimonadetes* (Table 1). Interestingly, we found that different isozymes encoded within the same genome were often phylogenetically divergent and categorized into distinct clades of complex I (see below), suggesting each isozyme may play a distinct physiological role in the cell. Supporting this view is our finding that many of the bacteria predicted to contain two complex I isozymes are known for their versatile energetic and physiological lifestyles, including *Rhodobacter sphaeroides* and *Rhodopseudomonas palustris* (aerobic and anaerobic respiration, anaerobic phototrophy) (25, 26), *Geobacter* species (different types of metal respiration) (27, 28), *Acidiphilium cryptum* (aerobic respiration, anoxic iron respiration) (29), and *Rhizobium etli* strains and *Sinorhizobium* strains (free-living and root nodule-associated lifestyles) (30) (Table 1). Although the physiological contribution of a second complex I isozyme is not clear, its association with organisms that are energetically and physiologically versatile is consistent with the hypothesis that individual isozymes may be important for different modes of growth.

There are five main clades of bacterial complex I. A phylogeny constructed from the concatenated amino acid sequences of the 14 complex I subunits (NuoA to NuoN) revealed five main clades, which we refer to as clades A to E (Fig. 2). These clades and the overall topology of the concatenated phylogeny are consistent with phylogenies of individual complex I subunits (see Fig. S1 to S7 in the supplemental material), indicating that the individual subunits share an evolutionary history and that recombination of complex I operons between different organisms is rare. As expected, complex I enzymes that did not fall into one of the five main clades in the concatenated phylogeny (Fig. 2, “Other”) were found to have individual complex I subunits that sorted into different parts of the phylogeny (either clade D or outside the five main clades) (see Fig. S1 to S7). These findings, taken in combination with the observation that genes encoding complex I are colocalized (likely part of a polycistronic operon) in 86% of com-

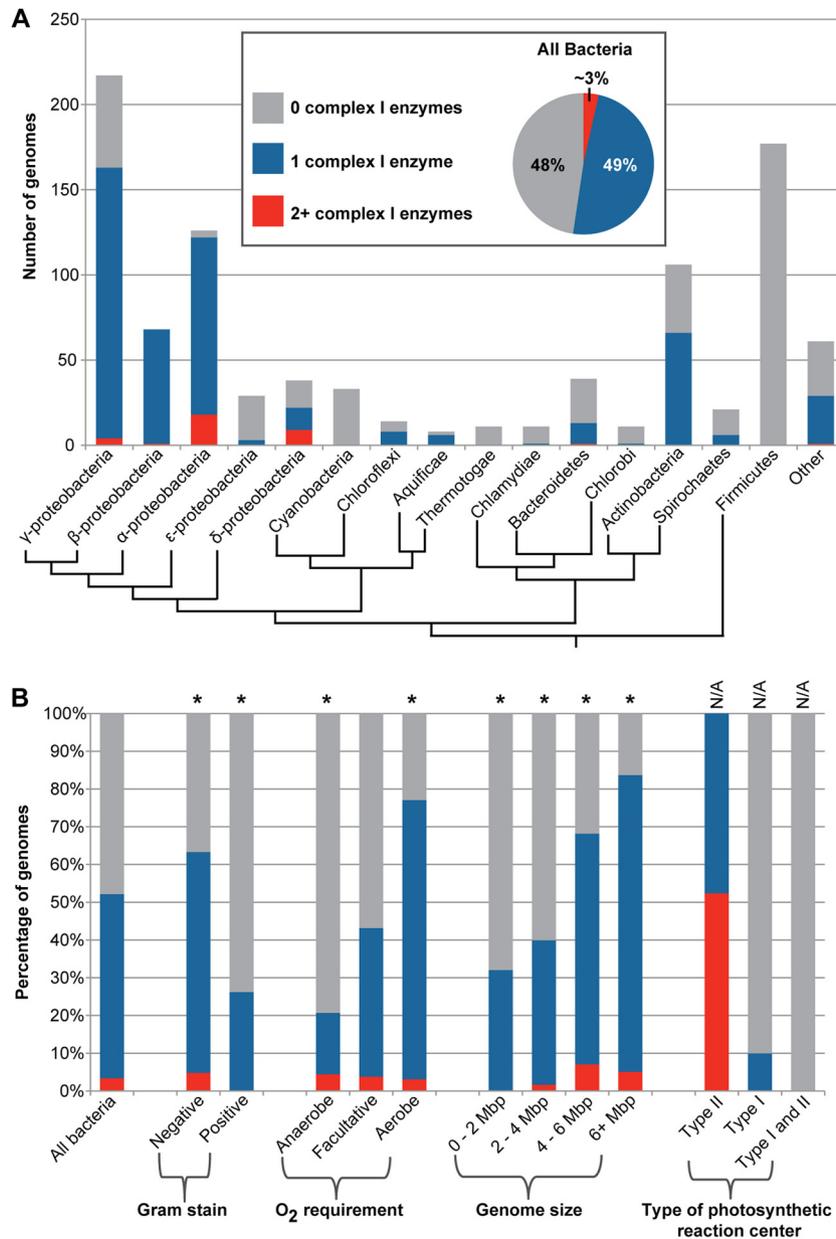


FIG 1 The distribution of complex I across bacteria. The distribution of predicted complex I enzymes in sequenced bacterial genomes. (A) The number of bacterial genomes predicted to encode 0, 1, or 2 complex I enzymes, organized by phylum or class. The *x* axis phylogeny was modeled using data from Ciccarelli et al. (77). (B) The percentage of bacterial genomes predicted to encode 0, 1, or 2 complex I enzymes, organized by trait. An asterisk (*) indicates that the distribution of complex I (0, 1, or 2 or more enzymes) for a specific trait is significantly different from the distribution of complex I in all bacteria (first bar) at *P* values of <0.0005 (chi-square test). "N/A" indicates there were not enough genomes in the category to perform the statistical analysis.

plex I-encoding genomes, suggest that these genes have overwhelmingly been vertically or horizontally transferred between bacterial lineages together rather than independently.

Complex I features were mapped onto the concatenated phylogeny to identify conserved and distinguishing characteristics between the five main enzyme clades (Fig. 2). For example, it has been noted that some bacterial complex I sequences have fused NuoC and NuoD subunits (31). Our analysis found that the presence of a fused NuoCD subunit was almost exclusively limited to clade E complex I. Other features that distinguished the different clade enzymes were the amino acid lengths of the NuoE and NuoG

subunits (involved in NADH binding and oxidation). On average, clade A enzymes had much longer NuoE subunits (262 amino acids), ranging from 166 to 457 amino acids in length. Generally, all complex I NuoE sequences are well conserved at the N terminus (the first ~170 amino acids), even between NuoE subunits from distantly related enzymes (Fig. 3, clade E NuoE from *E. coli* and clade A NuoE sequences). However, most clade A NuoE sequences have an extended C terminus, which is poorly conserved, even between closely related organisms (Fig. 3, clade A NuoE from *Rhodobacter sphaeroides* and *Rhodobacter capsulatus*). Although no potential functional significance can be assigned to the

TABLE 1 Genomes predicted to encode two or more complex I isozymes

| Phylum or class (no. of strains) | Strain | Enzyme clades | |
|--|---|--|------|
| Alphaproteobacteria (18) | <i>Acidiphilium cryptum</i> JF-5 | A, E | |
| | <i>Gluconacetobacter diazotrophicus</i> PAI 5 | A, E | |
| | <i>Rhizobium etli</i> (2 strains) | A, other | |
| | <i>Rhodobacter sphaeroides</i> (4 strains) | A, E | |
| | <i>Rhodospseudomonas palustris</i> (7 strains) | A, E | |
| | <i>Sinorhizobium fredii</i> NGR234 | A, other | |
| | <i>Sinorhizobium medicae</i> WSM419 | A, other | |
| | <i>Sinorhizobium meliloti</i> 1021 | A, other | |
| | <i>Chitinophaga pinensis</i> DSM 2588 | C, E | |
| | <i>Nitrosospora multififormis</i> ATCC 25196 | B, E | |
| | <i>Geobacter bemidjiensis</i> Bem | E, other | |
| | <i>Geobacter metallireducens</i> GS-15 | E, other | |
| | <i>Geobacter</i> sp. FRC-32 | E, other | |
| | <i>Geobacter</i> sp. M18 | E, other | |
| Bacteroidetes (1) Betaproteobacteria (1) Deltaproteobacteria (9) | <i>Geobacter</i> sp. M21 | E, other | |
| | <i>Geobacter sulfurreducens</i> PCA | E, other | |
| | <i>Geobacter uraniiireducens</i> Rf4 | E, other | |
| | <i>Pelobacter propionicus</i> DSM 2379 ^a | C, C, C | |
| | <i>Syntrophobacter fumaroxidans</i> MPOB ^b | E, E | |
| | Gammaproteobacteria (4) | <i>Gammaproteobacterium</i> HdN1 | E, E |
| | | <i>Nitrosococcus halophilus</i> Nc4 | B, E |
| | | <i>Nitrosococcus oceani</i> ATCC 19707 | B, E |
| | | <i>Nitrosococcus watsoni</i> C-113 | B, E |
| | Gemmatimonadetes (1) | <i>Gemmatimonas aurantiaca</i> T-27 | C, C |

^a The genome of this strain encoded three identical copies of operons predicted to encode complex I.

^b The genome of this strain was predicted to encode one copy of *nucEFG* genes and two copies of the remaining genes encoding complex I. Two isozymes could potentially be made from these components.

C-terminal tail of clade A NuoE proteins, work in *R. capsulatus* has shown that the NuoE C terminus is cleaved during anaerobic phototrophic growth conditions and particularly during diazotrophic conditions (32). The lengths of the NuoG subunits also varied greatly between different enzyme clades, where clade C enzymes had the shortest NuoG subunits and clade E enzymes had the longest, averaging 493 and 894 amino acids, respectively. Lastly, it was previously observed that some complex I NuoG subunits have a bacterium-specific iron-sulfur cluster N7, bound to a CXXXCXXC-(X)₂₇-C motif (33, 34). Our analysis found the N7 binding motif to be absent in most clade A and B enzymes but common in clade D and E enzymes, which also contain the longest NuoG subunits. In the clade E-containing *E. coli* enzyme, the N7 iron-sulfur cluster was shown to be important for enzyme stability but not involved in electron transfer (35), so this metal center is potentially required to stabilize larger NuoG subunits.

The concatenated complex I phylogeny generally followed the species phylogeny of the genomes in which it was encoded, although there were several notable exceptions. For example, though most gammaproteobacterial genomes were found to encode one complex I, the enzyme was one of two types: a clade B or a clade E enzyme (Fig. 2, red branches). Gammaproteobacteria with a clade B enzyme more closely follow the species phylogeny, as clade B also contains enzymes from the closely related betaproteobacteria. The observation that different members of gammaproteobacteria had only one of two distantly related complex I enzymes raises the possibility that each enzyme is important for a specific lifestyle (discussed below). Deltaproteobacterial genomes also encoded enzymes from different clades, including clades C and E and those that did not fall into one of the five main clades (Fig. 2, orange branches), although the relatively few genomes available from this group obviated a more complete investigation

of the potential physiological roles of the different clades of complex I in the deltaproteobacteria.

The complex I phylogeny also showed that the majority of bacteria with two complex I isozymes encoded enzymes from different clades (Table 1). In many of these genomes, we found that the phylogenetic placement of one complex I isozyme generally mirrored that of the bacterium in which it was encoded, while the other was apparently the product of an HGT event. For example, all alphaproteobacteria predicted to have (one or more) complex I encode a clade A enzyme, and the few alphaproteobacteria containing a second isozyme were predicted to encode either a clade E enzyme or an enzyme that did not fall into one of the five main clades. This suggests that clade A represents the canonical alphaproteobacterial complex I. Similar results were seen for other groups of bacteria, where the canonical betaproteobacterial and Bacteroidetes enzymes were found to be clades B and C, respectively. However, one betaproteobacterial (*Nitrosospora multififormis*) and one Bacteroidetes (*Chitinophaga pinensis*) genome were predicted to encode a clade E isozyme in addition to their canonical enzyme.

Gene loss and HGT determined the inheritance of complex I across the bacterial phylogeny. The distribution of complex I in gammaproteobacteria was unique in that some members did not have complex I, some members had a clade B complex I, others had a clade E enzyme, and a small number (all from the genus *Nitrosococcus*) had both. To better understand this inheritance pattern, we generated a species phylogeny of representative gammaproteobacteria from amino acid sequences of highly conserved single-copy housekeeping genes and mapped the predicted absence or presence of complex I onto each genome (Fig. 4). This analysis revealed a patchwork inheritance pattern of complex I in the gammaproteobacteria, suggesting that complex I genes have

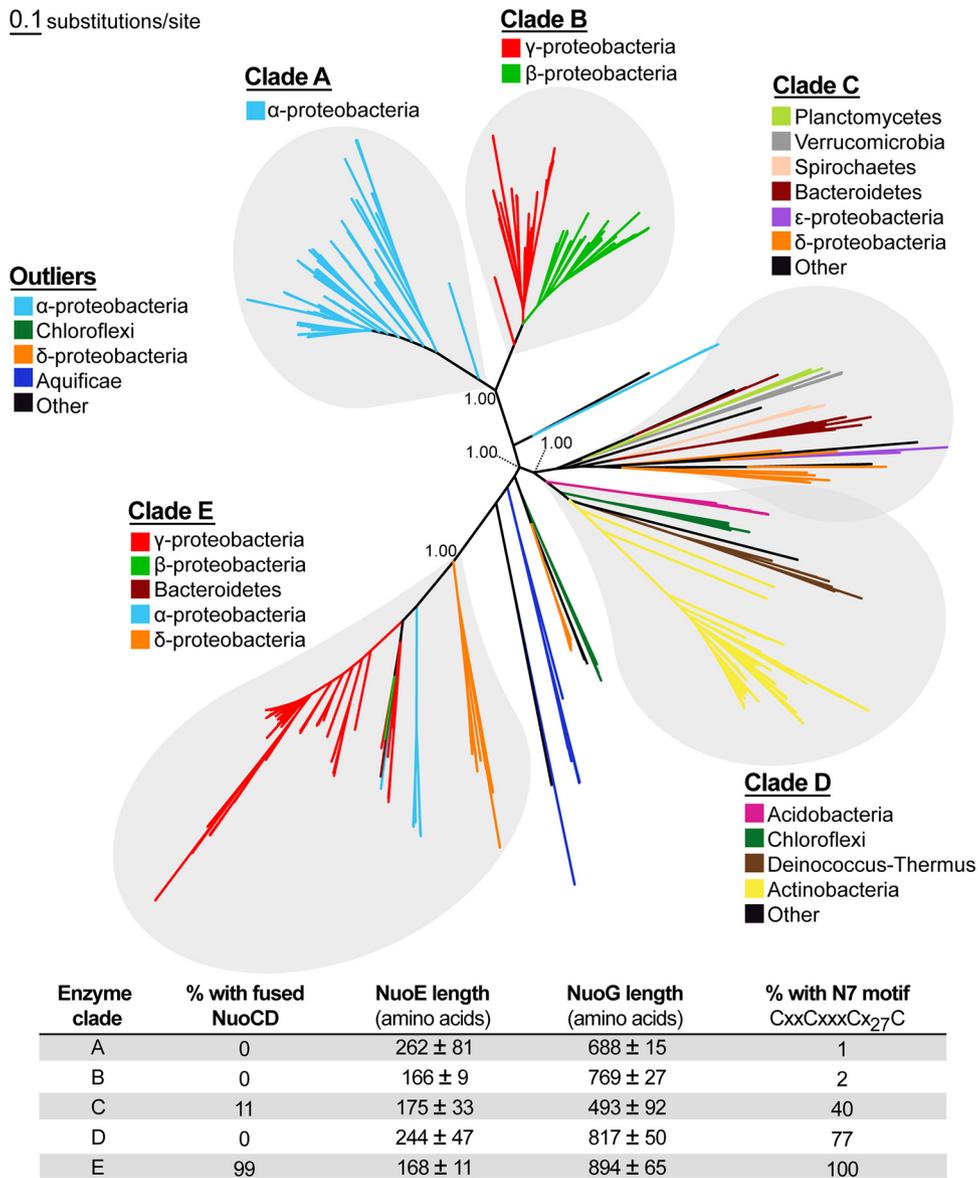


FIG 2 Phylogeny of bacterial complex I. A phylogeny of predicted complex I enzymes from 508 sequenced bacterial genomes was generated using amino acid sequences from all 14 concatenated complex I subunits. There are five main clades of bacterial complex I (clades A to E), which can be distinguished by specific subunit features. Genome names and most of the support values were omitted for clarity.

been lost and reacquired many times throughout the evolutionary history of this group. While most gammaproteobacteria that encode a clade B complex I were basal to those that encode a clade E enzyme, there were some exceptions. For example, *Methylococcus capsulatus* was predicted to encode a clade E enzyme (shaded purple in Fig. 4) despite being located in the basal portion of the gammaproteobacterial species tree, while the opposite pattern was observed for *Kangiella koreensis* (yellow in Fig. 4). Additionally, *Nitrosococcus* species were the only gammaproteobacteria predicted to have genes encoding both a clade B and clade E complex I, and these species were located at the interface of the clade B- and clade E-encoding gammaproteobacteria. This suggests that the initial introduction of a clade E enzyme to the gammaproteobacteria may have occurred around the time of the divergence of *Nitrosococcus*. This could potentially have led to the observed in-

heritance pattern, where the gammaproteobacterial species phylogeny consists of a derived group predominantly encoding a clade E complex I (purple in Fig. 4) and a basal group predominantly encoding a clade B enzyme (yellow in Fig. 4). Under this model, *Methylococcus capsulatus* and *Kangiella koreensis* would have lost and subsequently reacquired complex I.

The presence or absence of complex I was also mapped onto representative species phylogenies of other bacterial groups in which a large portion of members either had or lacked complex I, including the *Actinobacteria*, *Bacteroidetes*, and *Deltaproteobacteria* (Fig. 4). This analysis also showed a patchwork inheritance pattern of complex I within each of these groups, indicating that genes encoding the enzyme were lost and reacquired throughout their evolutionary history. Notably, many of the genomes predicted to encode two complex I isozymes (including the genomes

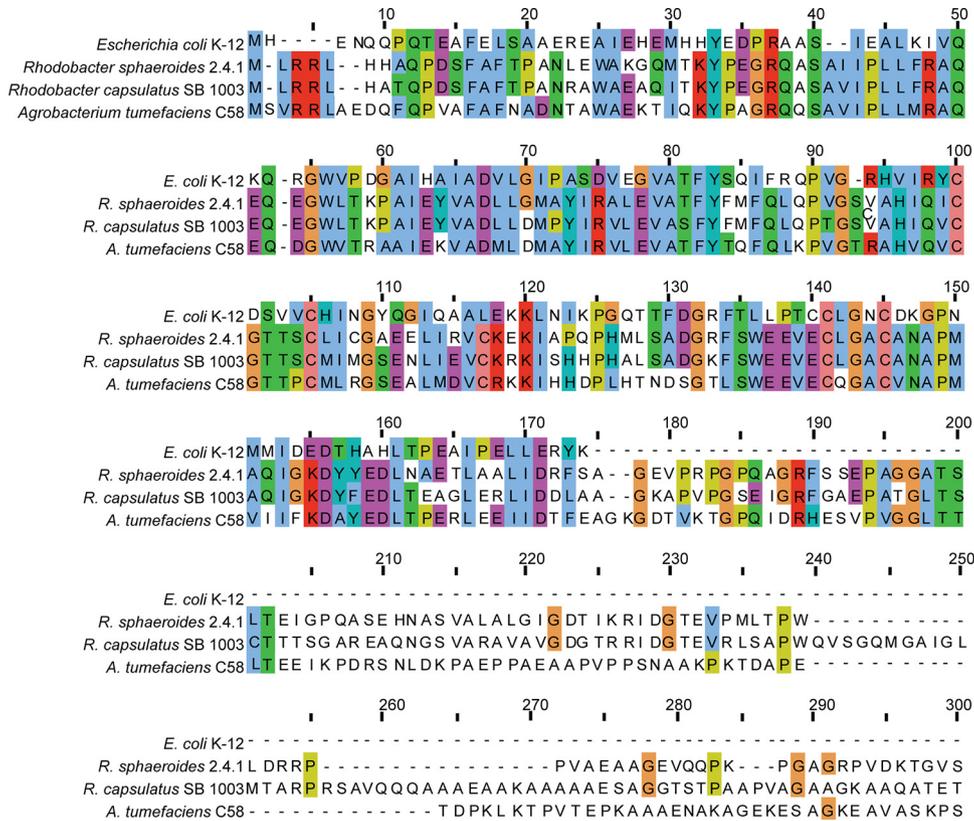


FIG 3 The extended C terminus of clade A NuoE is poorly conserved. Protein alignment of the first 300 amino acids of the complex I subunit NuoE from clade E *E. coli* and clade A *R. sphaeroides*, *R. capsulatus*, and *Agrobacterium tumefaciens*. The first ~170 amino acids of NuoE are well conserved between all organisms. Most clade A NuoE subunits have an extended C terminus, which is poorly conserved even between closely related bacteria.

of members of *Alphaproteobacteria*, *Betaproteobacteria*, *Deltaproteobacteria*, and *Bacteroidetes*) were found to encode a noncanonical enzyme that is phylogenetically similar to the *Nitrosococcus* clade E enzyme (Fig. 2). Because the clade E enzymes from these different bacteria are closely related to the *Nitrosococcus* clade E enzymes, it is likely that these clade E enzymes share an origin. Together, these observations support the hypothesis that the inheritance of complex I in these groups was largely the result of numerous gene losses and HGT events.

Complex I as a major determinant of NADH reoxidation and redox balance. In order to identify potential roles that complex I plays in bacterial physiology, we asked which biochemical pathways were present more often in bacterial genomes encoding complex I than in those lacking it within a particular phylum or class (Fig. 5). For example, of the 110 actinobacterial genomes analyzed in this study, 60 were predicted to encode complex I, while 44 were predicted to lack it. In this analysis, we compared all the annotated proteins belonging to actinobacterial genomes that encode complex I to all the annotated proteins in actinobacterial genomes that lack complex I. This provided us with a list of proteins and biochemical pathways that were more abundant, or “enriched” (Fisher’s exact test, $P < 0.05$), in actinobacterial genomes containing complex I than in those lacking the enzyme. As a positive control, we found complex I was enriched in the complex I-encoding bacteria compared to that in the complex I-lacking bacteria for each phylum or class analyzed, with the exception of the *Chloroflexi* (see Materials and Methods for more detail).

Approximately a third of the pathways that were enriched in complex I-encoding bacterial genomes are known or predicted to generate NADH as a product (Fig. 5). The association of this enzyme with NADH-generating pathways suggests that complex I is a respiratory chain entry point for electrons from these different pathways, linking catabolism to the generation of a PMF for ATP synthesis and other essential functions. It also suggests that complex I is important for reoxidizing the reduced pyridine nucleotide pool, thus maintaining a cellular redox state.

While many of the enriched pathways were specific to one phylum or class of bacteria, a number of central NADH-generating pathways were enriched across different groups of bacteria, including the full tricarboxylic acid (TCA) cycle, which produces 3 molecules of NADH with each complete cycle. Because the TCA cycle was enriched across different phyla, there may be a broad role for this enzyme to reoxidize NADH produced by the TCA cycle across bacteria. Likewise, this suggests that bacteria predicted to lack complex I use modified TCA cycles, and indeed this has been shown to be true in a number of bacteria lacking complex I, including *Cyanobacteria* (36), green sulfur bacteria (*Chlorobi*) (37), members of *Deltaproteobacteria* and *Epsilonproteobacteria* (37–41), Gram-positive bacteria (42, 43), and *Archaea* (44). In many cases, the function of these modified TCA cycles is still unclear. However, modified TCA cycles (e.g., branched TCA cycle) are thought to provide the cell with important biosynthetic precursors while producing less reductant (42, 45) or alternative types of reductant (NADPH instead of NADH) (36), which would

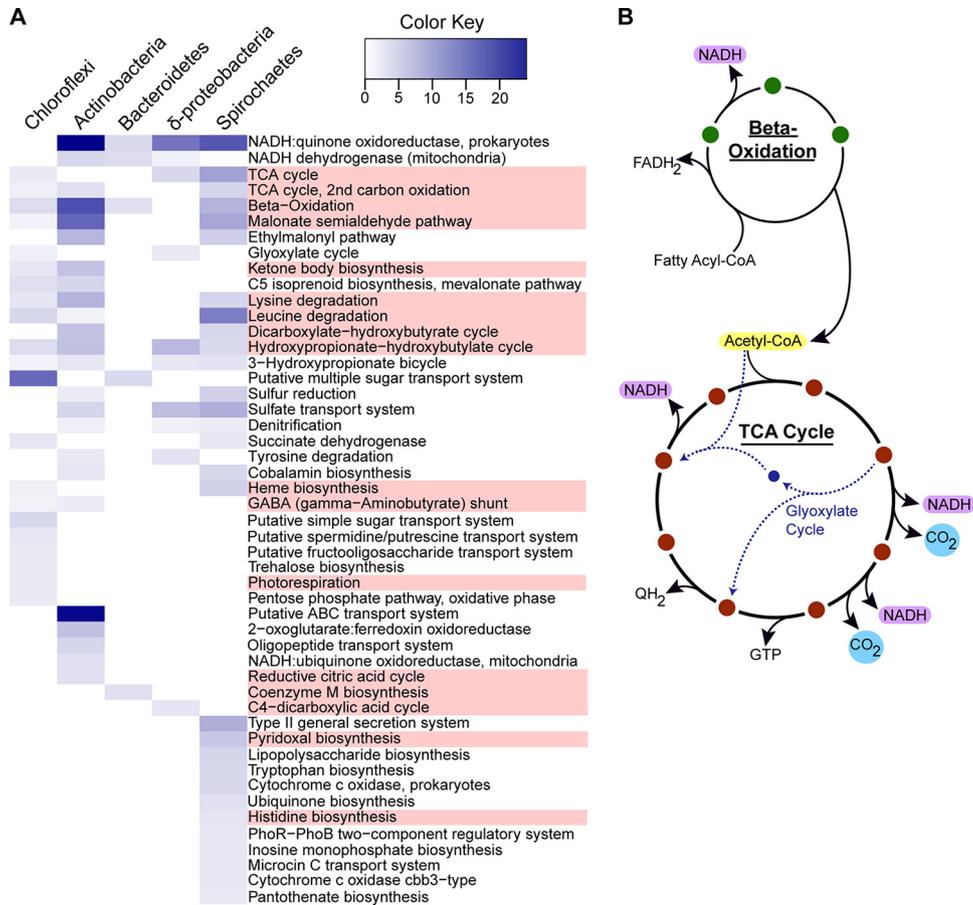


FIG 5 Biochemical pathways enriched in bacterial genomes that encode complex I. (A) Heat map showing the biochemical pathways (KEGG modules) that are enriched in bacterial genomes predicted to encode complex I compared to members of the same class/phylum that are predicted to lack complex I. We present pathways that show significant enrichment with a P value of <0.01 (see Data Set S2 in the supplemental materials for full results). Pathways are colored according to their level of significance, which was calculated by $-\log_{10}(P \text{ value})$, so that the most significantly enriched pathways have a higher value on the color key and are darker shades of blue. For visual purposes, the value of the most significant pathway (“NADH:quinone oxidoreductase, prokaryotes” in *Actinobacteria*) was adjusted from 44 to 25, which still represents the upper limit of the scale, while improving the resolution of other enriched pathways found in the analysis. Pathway names highlighted in pink produce NADH as a byproduct (72). (B) Schematic of biochemical pathways that are enriched across different groups of bacteria predicted to encode complex I. Beta-oxidation produces NADH and acetyl-CoA, the latter of which feeds into the TCA cycle to produce additional NADH. The glyoxylate cycle and ethylmalonyl pathway (not shown) assimilate acetyl-CoA by bypassing the CO_2 -generating reactions of the TCA cycle.

philic lithotroph *Acidithiobacillus ferrooxidans* and the halophilic purple sulfur bacterium *Halorhodospira halophila*. The clade E-containing gammaproteobacteria are mostly facultative anaerobes and include all of the enteric bacteria and *Pseudomonas* species used in this analysis. This observation provides a unique opportunity to understand the potential physiological implications of encoding a specific clade of complex I.

To identify potential physiological implications of encoding a specific clade of complex I, we tested for the enrichment of biochemical pathways in either clade B- or clade E-containing gammaproteobacteria compared to each other. Because clade E-containing gammaproteobacteria have, on average, larger genomes than clade B-containing gammaproteobacteria (~4,000 and ~2,800 annotated proteins per genome, respectively), statistical tests were performed between these groups within specific size ranges (e.g., genomes with 4,000 to 5,000 annotated proteins) to correct for pathways that might appear falsely enriched because they make up a larger fraction of the smaller genomes (i.e., essential functions). However, the difference in average genome sizes

between clade B- and clade E-containing gammaproteobacteria is itself an interesting observation, as many of the clade B-containing gammaproteobacteria live intracellularly and include obligate intracellular pathogens (e.g., *Coxiella burnetii*), facultative intracellular pathogens (e.g., *Legionella pneumophila*), and obligate intracellular mutualists (e.g., “*Candidatus* Ruthia magnifica”). Obligate intracellular organisms often have reduced genomes compared to those of free-living or generalist organisms (49). While it is not clear why a particular clade of complex I might be better suited to either a niche-specific or generalist lifestyle, and this pattern may be the result of bias in sequencing efforts in the current sequenced genome collection, the results of our comparative genomic analyses offer further insight.

Looking at the biochemical pathways that were enriched in at least two of the three genome size groups used in the comparative genomic analyses revealed that clade B-containing gammaproteobacteria were enriched for enzyme complexes of the canonical mitochondrial respiratory chain, including the cytochrome bc_1 complex and cytochrome c oxidase (Table 2). Because many clade

TABLE 2 Most significantly enriched biochemical pathways when comparing clade B- to clade E-containing gammaproteobacteria

| Pathways enriched in clade B-containing gammaproteobacteria, $P < 0.05$ | Pathways enriched in clade E-containing gammaproteobacteria, $P < 0.05$ |
|--|---|
| Cytochrome <i>bc</i> ₁ complex respiratory unit | Fumarate reductase, prokaryotes |
| Cytochrome <i>bc</i> ₁ complex | Dissimilatory nitrate reduction, nitrate → ammonia |
| Cytochrome <i>c</i> oxidase | Malonate semialdehyde pathway, propionyl-CoA → acetyl-CoA |
| Cytochrome <i>c</i> oxidase, prokaryotes | Valine/isoleucine biosynthesis, pyruvate → valine/2-oxobutanoate → isoleucine |
| RaxAB-RaxC type I secretion system | Isoleucine biosynthesis, threonine → 2-oxobutanoate → isoleucine |
| Type IV secretion system | Ascorbate degradation, ascorbate → D-xylulose-5P |
| <i>Helicobacter pylori</i> pathogenicity signature, <i>cagA</i> pathogenicity island | Catechol ortho-cleavage, catechol → 3-oxoadipate |
| HydH-HydG (metal tolerance) two-component regulatory system | Type III secretion system |
| CheA-CheYBV (chemotaxis) two-component regulatory system | Enterohemorrhagic/enteropathogenic <i>E. coli</i> pathogenicity signature, T3SS and effectors |
| Cph1-Rcp1 (light response) two-component regulatory system | RstB-RstA two-component regulatory system |
| PleC-PleD (cell fate control) two-component regulatory system | Sulfate transport system |
| Cysteine biosynthesis, homocysteine and serine → cysteine | Iron(III) transport system |
| Putative ABC transport system | Thiamine transport system |
| ABC-2-type transport system | Putative spermidine/putrescine transport system |
| Fatty acid biosynthesis, elongation | Glycine betaine/proline transport system |
| Ribosome, bacteria | L-Arabinose transport system |
| NADH:quinone oxidoreductase, prokaryotes ^a | Lysine/arginine/ornithine transport system |
| | Histidine transport system |
| | Glutamate/aspartate transport system |
| | General L-amino acid transport system |
| | Cystine transport system |
| | Branched-chain amino acid transport system |
| | Peptides/nickel transport system |
| | Iron complex transport system |
| | PTS system, glucose-specific II component |
| | PTS system, cellobiose-specific II component |
| | PTS system, ascorbate-specific II component |
| | Dipeptide transport system |
| | Microcin C transport system |
| | Taurine transport system |
| | Sulfonate transport system |
| | Oligopeptide transport system |

^a Only 2 of the 14 genes encoding complex I (*nuoC* and *nuoD*) were enriched in clade B-containing gammaproteobacteria. This is because these genes are separate in clade B operons (counted as 2 genes) and fused in clade E operons (counted as 0 genes). Thus, clade B-containing gammaproteobacteria appear to be enriched for complex I (NADH:quinone oxidoreductase, prokaryotes).

B-containing gammaproteobacteria are known pathogens, it was not surprising to also see enrichment for a number of virulence factors within this group, including type I and type IV secretion systems. In contrast, clade E-containing gammaproteobacteria were enriched for anaerobic respiratory enzymes, including fumarate and nitrate reductase (Table 2). Apart from anaerobic respiration, most of the other cellular functions that were enriched in clade E-containing gammaproteobacteria were transporters, which may indicate that these organisms live in variable environments where they are required to transport and metabolize a suite of nutrients (i.e., a more generalist lifestyle).

Previous analysis of complex I in *E. coli* supports the role of a clade E complex I enzyme in anaerobic fumarate respiration. In *E. coli*, complex I is not required for growth by aerobic respiration, but transcription of the genes encoding the enzyme is activated during nitrate respiration, and enzyme activity is essential for fumarate respiration (15, 50, 51). During aerobic respiration, *E. coli* utilizes a nonbioenergetic NADH dehydrogenase (NdhII) (15). It is thought that *E. coli* employs NdhII when sufficient ATP is generated through substrate-level phosphorylation and other enzymes (e.g., cytochrome *bo* oxidase) generate a PMF. Under these

conditions, NdhII allows the cell to maintain redox balance by oxidizing NADH without being sensitive to backpressure from the PMF (52). Conversely, *E. coli* has fewer systems for energy conservation during anaerobic respiration as compared to aerobic respiration, so the bioenergetic complex I is required for generating a PMF (15).

The insights gained from analyzing complex I function in *E. coli* may hold true in other bacteria. For example, bacteria that can metabolize an array of nutrients (clade E-containing gammaproteobacteria, generalists) may use bioenergetic complex I only when energy-generating options are limited (e.g., anaerobic respiration), in order to increase their opportunities for energy conservation. However, the association of complex I with aerobic respiratory enzymes in clade B-containing gammaproteobacteria suggests that these bacteria use complex I as a coupling site during aerobic respiration, possibly because these organisms live in nutrient-limited environments (e.g., low nutrient or O₂ levels). Future laboratory studies are needed to determine when the energetic contribution of complex I is important in various bacteria or to determine if clade B or E enzymes are better suited to aerobic or anaerobic respiratory conditions, respectively.

Physiological implications of bacterial complex I. This study found homologues of the 14-subunit mitochondrion-like proton-translocating NADH:quinone oxidoreductase (complex I) to be widespread in a large set of genomes from diverse bacteria. Though we predict its presence in bacteria with very different respiratory chains or energetic lifestyles, our findings suggest that a common role for complex I across bacterial phyla is to reoxidize NADH, thus maintaining the redox state of the cell. Therefore, we would predict bacteria without complex I (including the majority of Gram-positive bacteria) to have other enzymes or strategies for balancing the pyridine nucleotide pool, including fermentation, other types of NADH:quinone oxidoreductases (Ndh, Nqr) (53), and alternative NADH oxidases (NOXase, NPXase) (54–56), or to generate less NADH during catabolism, such as by using a modified TCA cycle (36, 42, 45). We used a comparative genomics approach to identify biochemical pathways that were enriched in bacteria predicted to lack complex I, but few trends emerged, suggesting that there are diverse strategies for growth without complex I. One trend that emerged from this analysis was the enrichment of enzymes that produce reduced ferredoxin (e.g., 2-oxoglutarate:ferredoxin oxidoreductase, pyruvate:ferredoxin oxidoreductase) in genomes lacking complex I. Perhaps one major strategy for bacteria without complex I is to use alternative forms of reductant (e.g., reduced ferredoxin), which can be reoxidized in subsequent metabolic pathways.

Our data also predict a varied role for complex I in bacteria that use light as an energy source, as only a subset of phototrophs were predicted to contain this enzyme. Specifically, none of the *Cyanobacteria* and only one member of the green sulfur bacteria (*Chlorobi*) but all of the purple bacteria and filamentous green nonsulfur bacteria (a subset of *Chloroflexi*) were predicted to encode complex I (Fig. 1B). The photochemical apparatuses of these groups differ in that the purple bacteria and filamentous green nonsulfur bacteria harbor only a type II reaction center, which uses membrane-bound quinones as electron acceptors (57). Thus, the presence of complex I in photosynthetic bacteria appears to be highly correlated with a photochemical system that produces quinol as a product of solar energy conservation. In this regard, it is interesting to note that previous work in the purple phototroph *Rhodobacter capsulatus* demonstrates that complex I is required for phototrophic growth, where the PMF drives complex I-mediated electron transfer from reduced quinone to NAD^+ , which prevents overreduction of the quinone pool and provides cellular reductant for biosynthetic purposes (16, 17). Our results support the hypothesis that complex I is required for solar energy conservation in all phototrophs with only a type II reaction center. In contrast, *Cyanobacteria* and green sulfur bacteria can potentially prevent overreduction of the quinone pool by using a type I reaction center to donate electrons to cytoplasmic ferredoxin, which can subsequently be used to synthesize NADPH for biosynthetic purposes (57). While complex I is known or predicted to synthesize NADH in nonphototrophic bacteria (e.g., chemolithotrophs), the extent to which the enzyme plays this role across bacteria is unknown (58, 59).

Our phylogenetic analyses showed that the majority of bacterial complex I enzymes could be placed into five main clades. In addition to their phylogenetic placement, enzymes within each clade could be distinguished based on subunit features (e.g., length of specific subunits, presence of the N7 iron-sulfur cluster), leading us to hypothesize that enzymes from different clades have

evolved to suit the individual lifestyles of their bacterial hosts. This hypothesis is supported by the observation that some closely related bacteria, such as the gammaproteobacteria, have distantly related complex I enzymes, which our pathway enrichment analysis suggests may be used in specific types of respiratory chains (aerobic and anaerobic). Further, a small number of bacterial genomes were predicted to encode genes for two complex I isozymes, usually from different enzyme clades. In the purple nonsulfur bacterium *Rhodobacter sphaeroides*, whole-genome expression studies show that its two complex I isozymes are differentially expressed in the presence and absence of oxygen (60), suggesting distinct physiological roles for these isozymes. Taken together, these results suggest the main clades of complex I have diverged physiologically as well as phylogenetically. Future genetic and biochemical studies must be performed to fully understand how the different clades of complex I contribute to individual bacterial lifestyles.

By comparing the phylogenies of individual complex I subunits to the concatenated (14-subunit) enzyme phylogeny, we demonstrate that subunit recombination between different clades of complex I is rare and that all subunits are generally lost or horizontally transferred together. Despite the pressure against hybrid enzymes, there are examples of modified complex I enzymes that have gained new function. Complex I has a modular design, where three subunits (NuoE, NuoF, NuoG) are involved in NADH binding and oxidation (61). Previous work has documented the abundance of partial complex I operons, which lack the three NADH-oxidizing subunits, across prokaryotes (23). Several experimental studies have shown that some microbes with partial complex I operons have incorporated other proteins into these partial enzyme complexes to create bioenergetic enzymes that utilize alternative electron donors. These microbes include *Campylobacter jejuni* (flavodoxin) (18), archaea (F_{420}H_2 , reduced ferredoxin) (19, 20), and *Cyanobacteria* (unknown electron donor) (21). Together, these findings suggest that partial complex I operons, rather than representing nonfunctional enzymes, have evolved novel physiological roles. Our own findings, which describe the different clades of bacterial complex I and their potential roles in physiology, lend support to the idea that complex I has been adapted in a variety of ways to suit the many physiological and energetic needs of the microorganisms in which it is found.

Conclusions. In this work, we show that homologues of mitochondrial complex I are widely distributed in bacteria. Complex I was found in bacteria with a range of energetic lifestyles, including obligate aerobes, obligate anaerobes, facultative anaerobes, obligate lithotrophs, and specific types of phototrophs (those with only a type II reaction center). We were also able to identify 5 main clades of bacterial complex I, with different members of closely related bacteria containing divergent complex I enzymes, while a small number of bacteria contained two complex I isozymes from different clades. In addition, our comparative genomic analyses predict that bacterial complex I plays a crucial role in oxidizing NADH from different catabolic reactions to serve as an entry point for electrons of the respiratory chain but that particular clades of enzymes may be adapted for specific physiological conditions (aerobic and anaerobic respiratory chains). Thus, results from this work predict complex I to have a more varied role in bacterial physiology than has been previously recognized. Our findings highlight the need for further studies to fully understand

the contribution of complex I to individual organisms and physiologies across the diversity of bacteria.

MATERIALS AND METHODS

Analysis of complex I occurrence across *Bacteria* and *Archaea*. Characterized bacterial complex I typically consists of 14 different protein subunits (NuoA to NuoN), which are considered the “minimal” subunits required for a functional enzyme (1, 62). To identify complex I gene clusters in prokaryotic genomes, we acquired 970 bacterial and 88 archaeal genomes and their associated protein predictions from NCBI (a full list of the genomes used is available in Data Set S1 in the supplemental material). Genomes were chosen to provide a broad sampling of prokaryotes, and genomes were excluded only if they were classified as a species for which multiple sequences of other strains were available (e.g., *E. coli*). The chosen genomes span 25 and 4 phyla of bacteria and archaea, respectively. Complex I protein subunit annotations were performed by comparing all of the proteins predicted to be encoded in the genomes to the NCBI Clusters of Orthologous Groups (COG) (63) database using RPS-BLAST (64) (E value of $1e-5$). Proteins were considered a predicted complex I subunit if their best RPS-BLAST hit was to a COG annotated as belonging to complex I (COG0838, COG0377, COG0852, COG0649, COG1905, COG1894, COG1034, COG1005, COG1143, COG0839, COG0713, COG1009, COG1008, and COG1007 for NuoA to NuoN, respectively).

Predicted complex I subunits and the chromosomal placement of their respective genes were manually inspected to confirm the presence of gene clusters encoding complex I within each genome. A genome was scored as encoding complex I if it was predicted to have genes for all 14 Nuo subunits (13 subunits if NuoC and NuoD were fused, 12 subunits if NuoB, NuoC, and NuoD were fused, etc.) typically found in characterized complex I enzymes. In cases where a genome was predicted to encode multiple copies of a single complex I subunit, chromosomal placement was used to determine which subunit was most likely part of complex I (since genes encoding the enzyme were typically colocalized). Putative complex I subunits that were located next to other putative complex I subunits were assumed to be part of complex I. Genomes missing one or more of the 14 highly conserved enzyme subunits (NuoA to NuoN) were scored as not encoding complex I. For genomes scored as encoding complex I, genes encoding complex I subunits (*nuaA* to *nunN*) were determined to be colocalized (potentially part of an operon) if less than 10 non-*nua* open reading frames (ORFs) separated any two genes of the operon.

Project metadata and data statistics for each available genome in the data set were acquired from the DOE Joint Genome Institute’s Integrated Microbial Genomes database. A complete list of the metadata used in this study is provided in Data Set S1 in the supplemental material. Chi-square tests were performed to determine if the distribution of complex I across all bacteria was different from the distribution of complex I for a given bacterial characteristic or lifestyle ($P < 0.0005$).

Phylogenetic tree construction. Proteins scored as being complex I subunits were compiled using custom PERL scripts, and phylogenetic trees were constructed from amino acid alignments of both individual complex I subunits and of the entire concatenated enzyme (14 subunits, NuoA to NuoN). Additionally, bacterial species trees were generated for genomes of interest using a concatenated alignment of the amino acid sequences of single-copy housekeeping genes: *recA*, *fusA*, *recG*, *rpoB*, *rplB*, *lepA*, *ileS*, *pyrG*, and *leuS*. These 9 genes have previously been shown to yield accurate phylogenetic trees of the genomes in which they are encoded (65). Hidden Markov models (HMMs) were created from curated alignments of these proteins available from the Ribosomal Database Project (66) by using the hmmbuild command in HMMER3 (67). Proteins encoded in the genomes of interest were compared to the HMMs using the hmmersearch command in HMMER3, and only best matches were retained. For all phylogenetic trees, alignments were created using MAFFT v. 5.662 E-INS-I (68). Phylogenetic trees were constructed using

FastTree (69), and support values for the nodes were calculated using the Shimodaira-Hasegawa (SH) test (70). Trees were visualized using the Interactive Tree of Life (iTOL) (71). Genome names and support values were omitted from figures for clarity, as necessary. Phylogenies of individual complex I subunits are available in Fig. S1 to S7 in the supplemental material. The full species phylogeny (including support values) that was generated from conserved housekeeping genes is available in Fig. S8 in the supplemental material.

Comparative genomic and bacterial physiology analyses. Comparative genomics was used to determine which biochemical pathways were enriched in bacterial genomes that encode complex I compared to bacterial genomes within the same phylum or class that lack the genes for complex I. This procedure was also used to determine which pathways were enriched in gammaproteobacteria containing a clade B complex I compared to gammaproteobacteria containing a clade E complex I (and vice versa). Here, genomes were annotated by uploading all predicted proteins to the Kyoto Encyclopedia of Genes and Genomes Automated Annotation Server (KEGG-KAAS) (72, 73). Annotations were parsed at both the enzyme and module levels as delineated on the KEGG website. Right- and left-tailed Fisher’s exact tests were used to test for relative enrichment and depletion of enzymes and modules between sets of genomes, and the resulting P values were corrected for multiple testing using the Benjamini-Hochberg correction (74). Modules and proteins having a P value of < 0.05 were considered significantly enriched. A complete list of enriched KEGG modules is provided in Data Set S2 in the supplemental material.

This analysis was also repeated within specific genome size ranges (e.g., genomes with 4,000 to 5,000 annotated proteins), because when the average number of encoded proteins differed significantly between the groups being compared, it resulted in the apparent enrichment of essential functions (e.g., ribosomal assembly) that are enriched only because they make up a larger fraction of a smaller genome. Using distinct genome size ranges did not change the results of comparing closely related genomes that have or lack genes for complex I, but it did alter the results of the analysis of the clade B- versus clade E-containing gammaproteobacteria. Thus, this analysis was performed between clade B- and clade E-containing gammaproteobacteria within specific genome size ranges: 2,000 to 3,000, 3,000 to 4,000, and 4,000 to 5,000 annotated proteins per genome.

It should be noted that when identifying biochemical pathways that were enriched in bacterial genomes containing complex I compared to genomes within the same phylum or class that lacked complex I, we looked for the enrichment of complex I (NADH:quinone oxidoreductase in KEGG) as a positive control. This analysis was able to identify the enrichment of complex I in the complex I-containing bacteria for each phylum or class analyzed, with the exception of the *Chloroflexi*. This exception likely reflects the fact that many *Chloroflexi* genomes that are predicted to lack complex I had genes encoding 13 of the 14 complex I subunits (missing only NuoG, which is required for a functional enzyme [75, 76]), while many genomes lacking complex I in the other phyla/classes used in this analysis were missing most of the complex I subunits. The statistical method used here was unable to distinguish between *Chloroflexi* genomes predicted to have complex I (14 subunits) and *Chloroflexi* genomes predicted to lack complex I that still had most of the genes required to encode the enzyme (13 subunits, lacking NuoG). Thus, complex I did not appear enriched when comparing *Chloroflexi* genomes with complex I to those lacking it.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00389-15/-/DCSupplemental>.

Figure S1, PDF file, 2.4 MB.

Figure S2, PDF file, 2.3 MB.

Figure S3, PDF file, 2.1 MB.

Figure S4, PDF file, 2.4 MB.

Figure S5, PDF file, 2.3 MB.

Figure S6, PDF file, 2.1 MB.
 Figure S7, PDF file, 2.5 MB.
 Figure S8, PDF file, 1.7 MB.
 Data Set S1, XLSX file, 0.1 MB.
 Data Set S2, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

This work was supported by the Department of Energy Office of Science's Great Lakes Bioenergy Research Center, grant DE-FC02-07ER64494. M.A.S. was supported by a traineeship from the NIGMS Biotechnology Training Program, grant T32 GM08349.

REFERENCES

- Brandt U. 2006. Energy converting NADH:quinone oxidoreductase (complex I). *Annu Rev Biochem* 75:69–92. <http://dx.doi.org/10.1146/annurev.biochem.75.103004.142539>.
- Sazanov LA. 2014. The mechanism of coupling between electron transfer and proton translocation in respiratory complex I. *J Bioenerg Biomembr* 46:247–253. <http://dx.doi.org/10.1007/s10863-014-9554-z>.
- Janssen RJ, Nijtmans LG, van den Heuvel LP, Smeitink JA. 2006. Mitochondrial complex I: structure and pathology. *J Inher Metab Dis* 29:499–515. <http://dx.doi.org/10.1007/s10545-006-0362-4>.
- Kussmaul L, Hirst J. 2006. The mechanism of superoxide production by NADH:ubiquinone oxidoreductase (complex I) from bovine heart mitochondria. *Proc Natl Acad Sci U S A* 103:7607–7612. <http://dx.doi.org/10.1073/pnas.0510977103>.
- Galkin A, Brandt U. 2005. Superoxide radical formation by pure complex I (NADH:ubiquinone oxidoreductase) from *Yarrowia lipolytica*. *J Biol Chem* 280:30129–30135. <http://dx.doi.org/10.1074/jbc.M504709200>.
- Fukui H, Moraes CT. 2008. The mitochondrial impairment, oxidative stress and neurodegeneration connection: reality or just an attractive hypothesis? *Trends Neurosci* 31:251–256. <http://dx.doi.org/10.1016/j.tins.2008.02.008>.
- Rhein V, Song X, Wiesner A, Ittner LM, Baysang G, Meier F, Ozmen L, Bluethmann H, Dröse S, Brandt U, Savaskan E, Czech C, Götz J, Eckert A. 2009. Amyloid-beta and tau synergistically impair the oxidative phosphorylation system in triple transgenic Alzheimer's disease mice. *Proc Natl Acad Sci U S A* 106:20057–20062. <http://dx.doi.org/10.1073/pnas.0905529106>.
- Lin MT, Beal MF. 2006. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature* 443:787–795. <http://dx.doi.org/10.1038/nature05292>.
- Richardson DJ. 2000. Bacterial respiration: a flexible process for a changing environment. *Microbiology* 146:551–571.
- Carroll J, Fearnley IM, Skehel JM, Shannon RJ, Hirst J, Walker JE. 2006. Bovine complex I is a complex of 45 different subunits. *J Biol Chem* 281:32724–32727. <http://dx.doi.org/10.1074/jbc.M607135200>.
- Baradaran R, Berrisford JM, Minhas GS, Sazanov LA. 2013. Crystal structure of the entire respiratory complex I. *Nature* 494:443–448. <http://dx.doi.org/10.1038/nature11871>.
- Torres-Bacete J, Sinha PK, Matsuno-Yagi A, Yagi T. 2011. Structural contribution of C-terminal segments of NuoL (ND5) and NuoM (ND4) subunits of complex I from *Escherichia coli*. *J Biol Chem* 286:34007–34014. <http://dx.doi.org/10.1074/jbc.M111.260968>.
- Sato M, Sinha PK, Torres-Bacete J, Matsuno-Yagi A, Yagi T. 2013. Energy transducing roles of antiporter-like subunits in *Escherichia coli* NDH-1 with main focus on subunit NuoN (ND2). *J Biol Chem* 288:24705–24716. <http://dx.doi.org/10.1074/jbc.M113.482968>.
- Grivennikova VG, Roth R, Zakharova NV, Hägerhäll C, Vinogradov AD. 2003. The mitochondrial and prokaryotic proton-trans locating NADH: ubiquinone oxidoreductases: similarities and dissimilarities of the quinone-junction sites. *Biochim Biophys Acta* 1607:79–90. <http://dx.doi.org/10.1016/j.bbabi.2003.09.001>.
- Tran QH, Bongaerts J, Vlad D, Uden G. 1997. Requirement for the proton-pumping NADH dehydrogenase I of *Escherichia coli* in respiration of NADH to fumarate and its bioenergetic implications. *Eur J Biochem* 244:155–160. <http://dx.doi.org/10.1111/j.1432-1033.1997.00155.x>.
- Dupuis A, Chevallet M, Darrouzet E, Duborjal H, Lunardi J, Issartel JP. 1998. The complex I from *Rhodobacter capsulatus*. *Biochim Biophys Acta* 1364:147–165. [http://dx.doi.org/10.1016/S0005-2728\(98\)00025-5](http://dx.doi.org/10.1016/S0005-2728(98)00025-5).
- Tichi MA, Meijer WG, Tabita FR. 2001. Complex I and its involvement in redox homeostasis and carbon and nitrogen metabolism in *Rhodobacter capsulatus*. *J Bacteriol* 183:7285–7294. <http://dx.doi.org/10.1128/JB.183.24.7285-7294.2001>.
- Weerakoon DR, Olson JW. 2008. The *Campylobacter jejuni* NADH: ubiquinone oxidoreductase (complex I) utilizes flavodoxin rather than NADH. *J Bacteriol* 190:915–925. <http://dx.doi.org/10.1128/JB.01647-07>.
- Baumer S, Ide T, Jacobi C, Johann A, Gottschalk G, Deppenmeier U. 2000. The F420H2 dehydrogenase from *Methanosarcina mazei* is a redox-driven proton pump closely related to NADH dehydrogenases. *J Biol Chem* 275:17968–17973. <http://dx.doi.org/10.1074/jbc.M000650200>.
- Welte C, Deppenmeier U. 2011. Membrane-bound electron transport in *Methanosaeta thermophila*. *J Bacteriol* 193:2868–2870. <http://dx.doi.org/10.1128/JB.00162-11>.
- Battchikova N, Eisenhut M, Aro EM. 2011. Cyanobacterial NDH-1 complexes: novel insights and remaining puzzles. *Biochim Biophys Acta* 1807:935–944. <http://dx.doi.org/10.1016/j.bbabi.2010.10.017>.
- Alexeeva S, Hellingwerf KJ, Teixeira de Mattos MJ. 2002. Quantitative assessment of oxygen availability: perceived aerobicity and its effect on flux distribution in the respiratory chain of *Escherichia coli*. *J Bacteriol* 184:1402–1406. <http://dx.doi.org/10.1128/JB.184.5.1402-1406.2002>.
- Moparthi VK, Hägerhäll C. 2011. The evolution of respiratory chain complex I from a smaller last common ancestor consisting of 11 protein subunits. *J Mol Evol* 72:484–497. <http://dx.doi.org/10.1007/s00239-011-9447-2>.
- Spehr V, Schlitt A, Scheide D, Guénebaut V, Friedrich T. 1999. Overexpression of the *Escherichia coli* nuo-operon and isolation of the overproduced NADH: ubiquinone oxidoreductase (complex I). *Biochemistry* 38:16261–16267. <http://dx.doi.org/10.1021/bi9919605>.
- Mackenzie C, Choudhary M, Larimer FW, Predki PF, Stilwagen S, Armitage JP, Barber RD, Donohue TJ, Hosler JP, Newman JE, Shapleigh JP, Sockett RE, Zeilstra-Ryalls J, Kaplan S. 2001. The home stretch, a first analysis of the nearly completed genome of *Rhodobacter sphaeroides* 2.4.1. *Photosynthesis Res* 70:19–41. <http://dx.doi.org/10.1023/A:1013831823701>.
- Larimer FW, Chain P, Hauser L, Lamerdin J, Malfatti S, Do L, Land ML, Pelletier DA, Beatty JT, Lang AS, Tabita FR, Gibson JL, Hanson TE, Bobst C, Torres JL, Peres C, Harrison FH, Gibson J, Harwood CS. 2004. Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodospseudomonas palustris*. *Nat Biotechnol* 22:55–61. <http://dx.doi.org/10.1038/nbt923>.
- Shelobolina ES, Vronis HA, Findlay RH, Lovley DR. 2008. *Geobacter uraniiireducens* sp. nov., isolated from subsurface sediment undergoing uranium bioremediation. *Int J Syst Evol Microbiol* 58:1075–1078. <http://dx.doi.org/10.1099/ijs.0.65377-0>.
- Lovley DR, Phillips EJ. 1988. Novel mode of microbial energy metabolism: organic carbon oxidation coupled to dissimilatory reduction of iron or manganese. *Appl Environ Microbiol* 54:1472–1480.
- Bilgin AA, Silverstein J, Jenkins JD. 2004. Iron respiration by *Acidiphilium cryptum* at pH 5. *FEMS Microbiol Ecol* 49:137–143. <http://dx.doi.org/10.1016/j.femsec.2003.08.018>.
- Becker A, Bergès H, Krol E, Bruand C, Rüberg S, Capela D, Lauber E, Meilhoc E, Ampe F, de Bruijn FJ, Fourment J, Francez-Charlot A, Kahn D, Küster H, Liebe C, Pühler A, Weidner S, Batut J. 2004. Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. *Mol Plant Microbe Interact* 17:292–303. <http://dx.doi.org/10.1094/MPMI.2004.17.3.292>.
- Friedrich T. 1998. The NADH:ubiquinone oxidoreductase (complex I) from *Escherichia coli*. *Biochim Biophys Acta* 1364:134–146. [http://dx.doi.org/10.1016/S0005-2728\(98\)00024-3](http://dx.doi.org/10.1016/S0005-2728(98)00024-3).
- Herter SM, Kortlüke CM, Drews G. 1998. Complex I of *Rhodobacter capsulatus* and its role in reverted electron transport. *Arch Microbiol* 169:98–105. <http://dx.doi.org/10.1007/s002030050548>.
- Sazanov LA, Hinchliffe P. 2006. Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*. *Science* 311:1430–1436. <http://dx.doi.org/10.1126/science.1123809>.
- Nakamaru-Ogiso E, Yano T, Yagi T, Ohnishi T. 2005. Characterization of the iron-sulfur cluster N7 (N1c) in the subunit NuoG of the proton-translocating NADH-quinone oxidoreductase from *Escherichia coli*. *J Biol Chem* 280:301–307. <http://dx.doi.org/10.1074/jbc.M410377200>.
- Pohl T, Bauer T, Dörner K, Stolpe S, Sell P, Zocher G, Friedrich T. 2007. Iron-sulfur cluster N7 of the NADH:ubiquinone oxidoreductase (complex I) is essential for stability but not involved in electron transfer. *Bio Chem* 46:6588–6596. <http://dx.doi.org/10.1021/bi700371c>.

36. Steinhäuser D, Fernie AR, Araújo WL. 2012. Unusual cyanobacterial TCA cycles: not broken just different. *Trends Plant Sci* 17:503–509. <http://dx.doi.org/10.1016/j.tplants.2012.05.005>.
37. Berg IA. 2011. Ecological aspects of the distribution of different autotrophic CO₂ fixation pathways. *Appl Environ Microbiol* 77:1925–1936. <http://dx.doi.org/10.1128/AEM.02473-10>.
38. Campbell BJ, Cary SC. 2004. Abundance of reverse tricarboxylic acid cycle genes in free-living microorganisms at deep-sea hydrothermal vents. *Appl Environ Microbiol* 70:6282–6289. <http://dx.doi.org/10.1128/AEM.70.10.6282-6289.2004>.
39. Callaghan AV, Morris BE, Pereira IA, McInerney MJ, Austin RN, Groves JT, Kukor JJ, Sufita JM, Young LY, Zylstra GJ, Wawrik B. 2012. The genome sequence of *Desulfatibacillum alkenivorans* AK-01: a blueprint for anaerobic alkane oxidation. *Environ Microbiol* 14:101–113. <http://dx.doi.org/10.1111/j.1462-2920.2011.02516.x>.
40. Heidelberg JF, Seshadri R, Haveman SA, Hemme CL, Paulsen IT, Kolonay JF, Eisen JA, Ward N, Methe B, Brinkac LM, Daugherty SC, Deboy RT, Dodson RJ, Durkin AS, Madupu R, Nelson WC, Sullivan SA, Fouts D, Haft DH, Selengut J, Peterson JD, Davidsen TM, Zafar N, Zhou LW, Radune D, Dimitrov G, Hance M, Tran K, Khouri H, Gill J, Utterback TR, Feldblyum TV, Wall JD, Voordouw G, Fraser CM. 2004. The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. *Nat Biotechnol* 22:554–559. <http://dx.doi.org/10.1038/nbt959>.
41. McInerney MJ, Rohlin L, Mouttaki H, Kim U, Krupp RS, Rios-Hernandez L, Sieber J, Struchtemeyer CG, Bhattacharyya A, Campbell JW, Gunsalus RP. 2007. The genome of *Syntrophus aciditrophicus*: life at the thermodynamic limit of microbial growth. *Proc Natl Acad Sci U S A* 104:7600–7605. <http://dx.doi.org/10.1073/pnas.0610456104>.
42. Amador-Noguez D, Feng XJ, Fan J, Roquet N, Rabitz H, Rabinowitz JD. 2010. Systems-level metabolic flux profiling elucidates a complete, bifurcated tricarboxylic acid cycle in *Clostridium acetobutylicum*. *J Bacteriol* 192:4452–4461. <http://dx.doi.org/10.1128/JB.00490-10>.
43. Morishita T, Yajima M. 1995. Incomplete operation of biosynthetic and bioenergetic functions of the citric acid cycle in multiple auxotrophic lactobacilli. *Biosci Biotechnol Biochem* 59:251–255. <http://dx.doi.org/10.1271/bbb.59.251>.
44. Danson MJ. 1989. Central metabolism of the archaeobacteria: an overview. *Can J Microbiol* 35:58–64. <http://dx.doi.org/10.1139/m89-009>.
45. Meléndez-Hevia E, Waddell TG, Cascante M. 1996. The puzzle of the Krebs citric acid cycle: assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution. *J Mol Evol* 43:293–303. <http://dx.doi.org/10.1007/BF02338838>.
46. Erb TJ, Berg IA, Brecht V, Müller M, Fuchs G, Alber BE. 2007. Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc Natl Acad Sci U S A* 104:10631–10636. <http://dx.doi.org/10.1073/pnas.0702791104>.
47. Chen GQ, Wu Q. 2005. Microbial production and applications of chiral hydroxyalkanoates. *Appl Microbiol Biotechnol* 67:592–599. <http://dx.doi.org/10.1007/s00253-005-1917-2>.
48. Steinbüchel A, Fuchtenbusch B. 1998. Bacterial and other biological systems for polyester production. *Trends Biotechnol* 16:419–427. [http://dx.doi.org/10.1016/S0167-7799\(98\)01194-9](http://dx.doi.org/10.1016/S0167-7799(98)01194-9).
49. McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10:13–26. <http://dx.doi.org/10.1038/nrmicro2670>.
50. Bongaerts J, Zoske S, Weidner U, Unden G. 1995. Transcriptional regulation of the proton-translocating NADH dehydrogenase genes (*nuoA-N*) of *Escherichia coli* by electron-acceptors, electron-donors and gene regulators. *Mol Microbiol* 16:521–534. <http://dx.doi.org/10.1111/j.1365-2958.1995.tb02416.x>.
51. Salmon K, Hung SP, Mekjian K, Baldi P, Hatfield GW, Gunsalus RP. 2003. Global gene expression profiling in *Escherichia coli* K-12. The effects of oxygen availability and FNR. *J Biol Chem* 278:29837–29855. <http://dx.doi.org/10.1074/jbc.M213060200>.
52. Bekker M, de Vries S, Ter Beek A, Hellingwerf KJ, de Mattos MJ. 2009. Respiration of *Escherichia coli* can be fully uncoupled via the nonelectrogenic terminal cytochrome *bd-II* oxidase. *J Bacteriol* 191:5510–5517. <http://dx.doi.org/10.1128/JB.00562-09>.
53. Kerscher S, Dröse S, Zickermann V, Brandt U. 2008. The three families of respiratory NADH dehydrogenases. *Results Probl Cell Differ* 45: 185–222. http://dx.doi.org/10.1007/400_2007_028.
54. Ross RP, Claiborne A. 1991. Cloning, sequence and overexpression of NADH peroxidase from *Streptococcus faecalis* 10C1—structural relationship with the flavoprotein disulfide reductases. *J Mol Biol* 221:857–871. [http://dx.doi.org/10.1016/0022-2836\(91\)80180-3](http://dx.doi.org/10.1016/0022-2836(91)80180-3).
55. Ross RP, Claiborne A. 1992. Molecular cloning and analysis of the gene encoding the NADH oxidase from *Streptococcus faecalis* 10C1. Comparison with NADH peroxidase and the flavoprotein disulfide reductases. *J Mol Biol* 227:658–671. [http://dx.doi.org/10.1016/0022-2836\(92\)90215-6](http://dx.doi.org/10.1016/0022-2836(92)90215-6).
56. Gibson CM, Mallett TC, Claiborne A, Caparon MG. 2000. Contribution of NADH oxidase to aerobic metabolism of *Streptococcus pyogenes*. *J Bacteriol* 182:448–455. <http://dx.doi.org/10.1128/JB.182.2.448-455.2000>.
57. Bryant DA, Frigaard NU. 2006. Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol* 14:488–496. <http://dx.doi.org/10.1016/j.tim.2006.09.001>.
58. Chain P, Lamerdin J, Larimer F, Regala W, Lao V, Land M, Hauser L, Hooper A, Klotz M, Norton J, Sayavedra-Soto L, Arciero D, Hommes N, Whittaker M, Arp D. 2003. Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. *J Bacteriol* 185:2759–2773. <http://dx.doi.org/10.1128/JB.185.21.6496.2003>.
59. Elbehti A, Brasseur G, Lemesle-Meunier D. 2000. First evidence for existence of an uphill electron transfer through the bc(1) and NADH-Q oxidoreductase complexes of the acidophilic obligate chemolithotrophic ferrous ion-oxidizing bacterium *Thiobacillus ferrooxidans*. *J Bacteriol* 182: 3602–3606. <http://dx.doi.org/10.1128/JB.182.12.3602-3606.2000>.
60. Arai H, Roh JH, Kaplan S. 2008. Transcriptome dynamics during the transition from anaerobic photosynthesis to aerobic respiration in *Rhodospirillum rubrum*. *J Bacteriol* 190:286–299. <http://dx.doi.org/10.1128/JB.01375-07>.
61. Friedrich T, Scheide D. 2000. The respiratory complex I of bacteria, archaea and eukarya and its module common with membrane-bound multisubunit hydrogenases. *FEBS Lett* 479:1–5. [http://dx.doi.org/10.1016/S0014-5793\(00\)01867-6](http://dx.doi.org/10.1016/S0014-5793(00)01867-6).
62. Sazanov LA. 2007. Respiratory complex I: mechanistic and structural insights provided by the crystal structure of the hydrophilic domain. *Bio Chem* 46:2275–2288. <http://dx.doi.org/10.1021/bi602508x>.
63. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. <http://dx.doi.org/10.1186/1471-2105-4-41>.
64. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <http://dx.doi.org/10.1093/nar/25.17.3389>.
65. Santos SR, Ochman H. 2004. Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environ Microbiol* 6:754–759. <http://dx.doi.org/10.1111/j.1462-2920.2004.00617.x>.
66. Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM. 2007. The ribosomal database project. RDP-II: introducing myRDP space and quality controlled public data. *Nucleic Acids Res* 35:D169–D172. <http://dx.doi.org/10.1093/nar/gkl889>.
67. Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference. *Genome Inform Int Conf Genome Inform* 23: 205–211.
68. Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33: 511–518. <http://dx.doi.org/10.1093/nar/gki198>.
69. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <http://dx.doi.org/10.1371/journal.pone.0009490>.
70. Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol* 55:539–552. <http://dx.doi.org/10.1080/10635150600755453>.
71. Letunic I, Bork P. 2007. Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127–128. <http://dx.doi.org/10.1093/bioinformatics/btl529>.
72. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36:D480–D484. <http://dx.doi.org/10.1093/nar/gkm882>.
73. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS:

- an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35:W182–W185. <http://dx.doi.org/10.1093/nar/gkm321>.
74. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc B Stat Methodol* 57:289–300.
75. Falk-Krzesinski HJ, Wolfe AJ. 1998. Genetic analysis of the *nuo* locus, which encodes the proton-translocating NADH dehydrogenase in *Escherichia coli*. *J Bacteriol* 180:1174–1184.
76. Erhardt H, Steimle S, Muders V, Pohl T, Walter J, Friedrich T. 2012. Disruption of individual *nuo*-genes leads to the formation of partially assembled NADH:ubiquinone oxidoreductase (complex I) in *Escherichia coli*. *Biochim Biophys Acta* 1817:863–871. <http://dx.doi.org/10.1016/j.bbabi.2011.10.008>.
77. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* 311:1283–1287. <http://dx.doi.org/10.1126/science.1123061>.