

Asymmetrical Evolution of Cytochrome *bd* Subunits

Weilong Hao, G. Brian Golding

Department of Biology, McMaster University, Hamilton, Ontario, Canada L8S 4K1

Received: 7 January 2005 / Accepted: 18 September 2005 [Reviewing Editor: Dr. Brian Morton]

Abstract. Functionally linked genes generally evolve at similar rates and the knowledge of this particular feature of genomic evolution has been used as the basis for the phylogenetic profiling method. We illustrate here an exception to this rule in the evolution of the cytochrome *bd* complex. This is a two-component oxidase complex, with the subunits I and II known to be widely present in bacteria. The subunits within the cytochrome *bd* complex are under the same evolutionary pressure and most likely behave in the same evolutionary manner. However, the sequence similarity of genes encoding subunit II varies considerably across species. Genes encoding subunit II evolve 1.2 times faster on most of the branches of their phylogeny than subunit I genes. Furthermore, the genes encoding subunit II in *Oceanobacillus iheyensis*, *Bacillus halodurans*, and *Staphylococcus* species do not have detectable homologues within *E. coli* due to their large divergence. Together, the two subunits of cytochrome *bd* reveal an interesting example of an asymmetric pattern of evolutionary change.

Key words: Asymmetrical evolution — Cytochrome *bd* — Phylogeny — Subunit

Introduction

Cytochrome *bd* is a terminal oxidase in the branched electron transport chain (Dassa et al. 1991). The en-

zyme complex consists of two distinct subunits; subunit I and subunit II. In *E. coli* K12, subunits I and II are encoded by the *appCB* operon or the *cydAB* operon. The gene *appC* encodes 514 amino acid residues and shares 61% sequence identity with *cydA*, which encodes 523 amino acid residues. Similarly, *appB* encodes 378 amino acid residues and shares 57% sequence identity with *cydB*, which encodes 377 amino acids (Blattner et al. 1997). These two operons might be functionally equivalent in *E. coli* (Trumpower and Gennis 1994; Sturr et al. 1996). Cytochrome *bd* complexes are widely known to be present in bacteria (Junemann 1997).

Subunit I has been predicted to contain nine transmembrane helices, while subunit II is predicted to have eight transmembrane helices (Osborne and Gennis 1999). In subunit I, there is a predicted loop near the outer (periplasmic) surface of the membrane connecting the eighth and ninth transmembrane helices called the Q-loop. Bacterial species can be classified into two groups according to the Q-loop region. One group contains the C-terminal half of the Q-loop while the other group is deleted for this region. The former group includes most of the γ -proteobacteria, while the latter group includes all of the Gram-positive bacteria, the cyanobacterium, and the γ -proteobacterium *Pseudomonas aeruginosa* (Osborne and Gennis 1999; Sakamoto et al. 1999). It has been demonstrated that the Q-loop is adjacent to loop I-II in subunit II (Ghaim et al. 1995). Furthermore, the Q-loop is known to be involved in quinol binding (Ghaim et al. 1995). The cytochrome *bd* complex contains three prosthetic groups: heme *b*₅₅₈, heme *b*₅₉₅, and heme *d*. Heme *b*₅₅₈ and heme *b*₅₉₅ are protoporphyrin, while heme *d* is a chlorin (Timkovich

Correspondence to: G. Brian Golding; email: golding@mcmaster.ca

et al. 1985). All three of the heme prosthetic groups are located on the periplasmic side of the membrane. Heme b_{558} is known to be located entirely within subunit I (Green et al. 1984). The Q-loop in subunit I has been suggested to participate in the binding of heme d and subunit II is necessary to bind heme b_{595} and heme d in *E. coli* (Green et al. 1984, 1986).

The two cytochrome *bd* subunits function as a whole, thus the association between the two subunits should be considered strong. It has been suggested that functionally linked proteins evolve in a correlated fashion (Pellegrini et al. 1999; Marcotte et al. 2000). Similarly, genes that are coexpressed show a similar rate of evolution (Jordan et al. 2004). This feature of molecular evolution is now being widely used as an aid in the annotation of genomes (Zheng et al. 2002; Gutierrez et al. 2004; Mikkelsen et al. 2005). The subunits of a metabolic enzyme are often under the same evolutionary pressure, and therefore, they generally behave in the same manner (Ciccarese et al. 1997). Since the subunits of the two-component complex cytochrome *bd* are most likely involved in similar evolutionary processes, they are expected to evolve at similar rates.

In contrast, lateral gene transfer (LGT) is known to cause apparent abrupt changes of evolutionary rate (Jain et al. 1999; Gogarten et al. 2002). Several recent studies have suggested that lateral gene transfer of a whole enzyme complex might have occurred (Lawrence and Ochman 1998; Osborne and Gennis 1999; Iyer et al. 2004). A possible example of this in the cytochrome *bd* complex genes is the phylogenetic clustering of the γ -proteobacterium *P. aeruginosa* gene complex with the α -proteobacteria gene complex (Osborne and Gennis 1999; Sakamoto et al. 1999). But the lateral transfer of a single subunit of a large macromolecular complex is comparatively rare due to the coevolution of these subunits and their interactions with other subunits. In part, this type of observation prompted Jain et al. (1999)'s suggestion of the "complexity hypothesis" of lateral gene transfer where the greater the number of interactions a gene has, the less likely it is to be involved in transfers.

This study was conducted to determine if the evolutionary rates between the two oxidase subunits are different and to explore how such differences might have contributed to their evolution. We show that the sequence similarity of subunit II shows an abrupt discontinuity between some species, while the sequence similarity of subunit I follows the inferred phylogenetic history without abrupt changes. An obvious lateral gene transfer of one subunit is not present, because the two subunits have similar phylogenies. Subunit II genes show an accelerated evolutionary pattern on most branches in comparison to the subunit I genes. The elevated rate of evolution of

subunit II on certain branches is more dramatic than that on other branches. These genes therefore demonstrate an interesting example of subunit evolution with asymmetrical rates of evolutionary change.

Methods

The protein sequences of *cydA* (subunit I) and *cydB* (subunit II) were extracted from the complete *E. coli* K12 genome sequence. The protein sequences of *cydA* and *cydB* in *E. coli* K12 were used as query sequences to search against 110 other whole bacterial genome sequences. Complete bacterial genome sequences were downloaded from NCBI (<http://www.ncbi.nlm.nih.gov/>). The NCBI numbers of these genomes are given as supplementary information at <http://life.biology.mcmaster.ca/~weilong/subunit>. The similarity of genomic proteins to the query proteins were measured via the BLASTP algorithm (Altschul et al. 1997). All hits with an expect value less than 10^{-5} were deemed to be potential homologues.

The best hits to *E. coli* K12's *cydA/cydB* in each species were extracted to conduct pairwise distance analyses. In general, cytochrome *bd* subunit genes have been found to exist within the same operon (Dassa et al. 1991; Trumpower and Gennis 1994; Sturr et al. 1996). Therefore, to ensure that the hits of *cydAB* are both orthologs of *cydAB* rather than potentially being misled by other close hits (Koski and Golding 2001), the physical locations of the hits on each chromosome were individually checked. Only the two gene homologues that are adjacently located on the chromosome were used for further pairwise distance analyses.

Multiple alignments of these sequences were constructed using the program CLUSTALW (Thompson et al. 1994). Pairwise distances between the genes from different species were measured by PUZZLE (Strimmer and von Haeseler 1996). The maximum likelihood method was employed to construct phylogenies with less than 20 taxa using the PHYLIP package (Felsenstein 1989) version 3.6 of 2004. Phylogenetic trees with a larger number of taxa were generated using PUZZLE (Strimmer and von Haeseler 1996) and using NEIGHBOR (Felsenstein 1989). Then the maximum likelihood method was used to compare the PUZZLE and NEIGHBOR trees. The tree with the higher likelihood was used for further consideration.

A broader survey was conducted to extract more potential homologues. In this case, 19 archaeal genomes were included with the 110 bacterial genomes (the NCBI numbers of archaeal genomes are also given at <http://life.biology.mcmaster.ca/~weilong/subunit>). Each of the significant hits of *cydAB* in *E. coli* as found above were then used as a query sequence in turn. All of the hits including duplicate subunit pairs in genome from these searches were collected and duplicates from the multiple searches eliminated to obtain a broad range of potential homologues. Again, all the genes were checked to determine if the potential homologues to subunits I and II were physically adjacent to each other in each genome.

A relative rate test was conducted using Tajima's (1993) method to test the difference in rates between each subunit. In addition, the distance-matrix rate test (Syvanen 2002) was employed using distances measured by PUZZLE and standardized to the number of replacements per 100 residues by multiplying 100.

Results

Based on sequence similarity to the *E. coli* K12's *cydA* gene, the database search revealed 76 bacterial genomes that have *cydA* (subunit I) homologues (shown in Fig. 1). The remaining 34 bacterial ge-

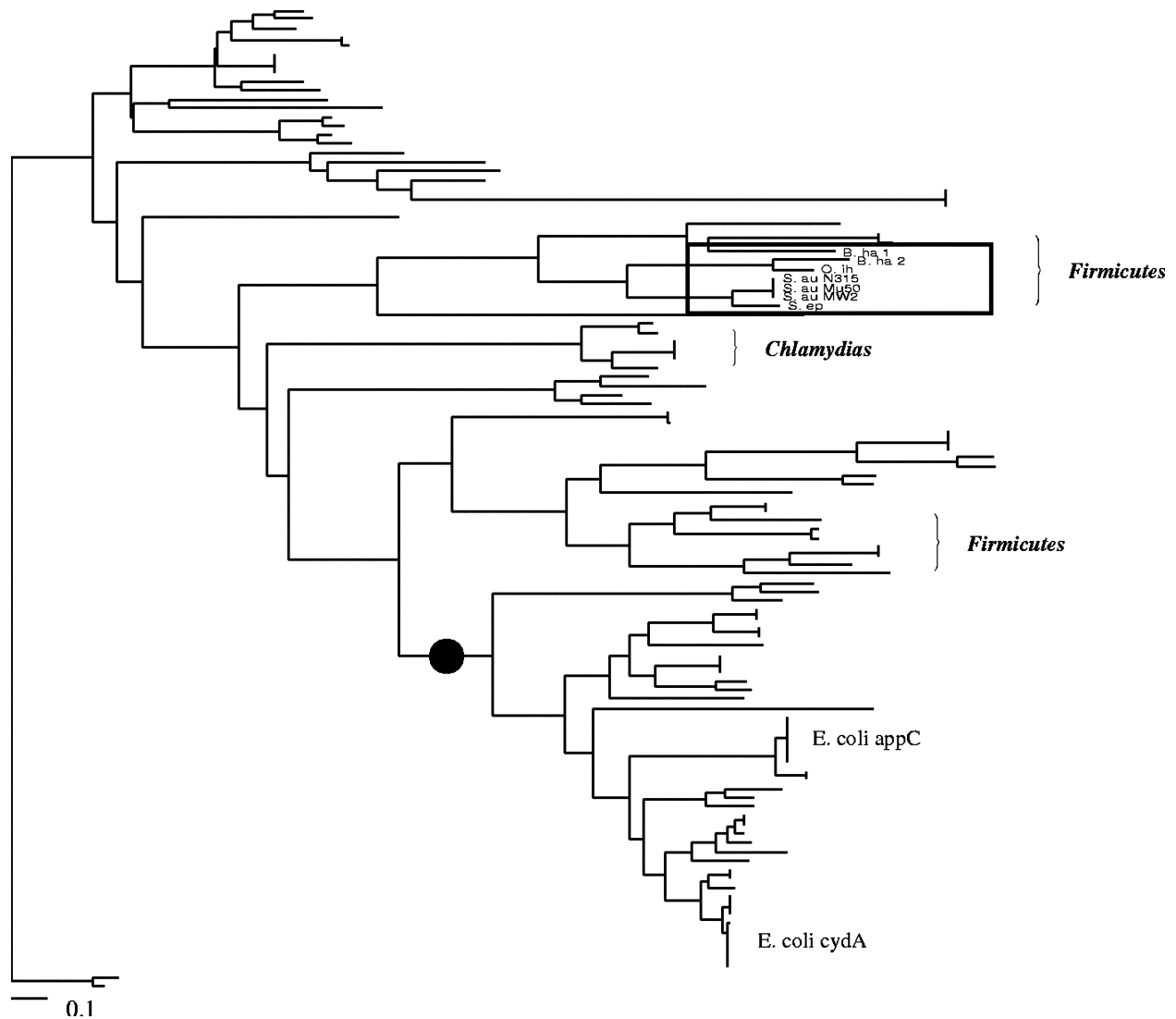


Fig. 1. Unrooted tree of subunit I genes. The genes shown here have expect values $< 10^{-5}$ with the *cydA* gene from *E. coli* using the BLASTP algorithm. While the six boxed species have homologues to *cydA*, they do not have *cydB* homologues. These six are

Staphylococcus epidermidis, *S. aureus* N315, *S. aureus* Mu50, *S. aureus* MW2, *Oceanobacillus iheyensis*, and *Bacillus halodurans* (which has two copies of the subunit I genes). The filled circle indicates the origin of the Q-loop cluster.

nomes searched do not have detectable homologues with any cytochrome *bd* subunit in *E. coli*. Of the 76 genomes, six genomes have only *cydA* homologues and no *cydB* (subunit II) homologues with *E. coli*. These six genomes are *Bacillus halodurans*, *Staphylococcus aureus* N315, *S. aureus* Mu50, *S. aureus* MW2, *Oceanobacillus iheyensis*, and *S. epidermidis*. The boxed region in Fig. 1 highlights these six species. However, subunit II genes have been found physically adjacent to subunit I genes in these six genomes according to the KEGG (Ogata et al. 1999) and EcoCyc databases (Karp et al. 2002). Yet the subunit II gene in *O. iheyensis*, for example, has no detectable similarity with *E. coli cydB* or *appB* (expect value greater than 0.1). The subunit II gene in *O. iheyensis* is, however, closely related with the genes in *B. halodurans*, *B. stearothermophilus*, *S. aureus*,

S. epidermidis, and *Exiguobacterium* sp. 255 (Table 1), but, the subunit II gene in *O. iheyensis* is more distantly related to the genes in other *Bacillus* species including *B. anthracis* and *B. cereus*, compared with the *Staphylococcus* species. The failure to detect the subunit II genes of these six genomes in the sequence similarity search, therefore, is not due to annotation error. The sequences for the subunit II genes from these six species were extracted from their complete genome sequences. The phylogenetic relationships of these subunit II sequences are shown in Fig. 2B. The phylogenetic relationships of the subunit I genes from these six species are shown in Fig. 2A for comparison.

The protein sequence of *O. iheyensis* subunit II sequence, when searched with BLASTP, revealed a very peculiar situation. According to the BLASTP

Table 1. The subunit II gene sequence from *Oceanobacillus iheyensis* was used as a query to BLAST against the NCBI database: All hits with an expect value $< 10^{-5}$ are listed

Query <i>Oceanobacillus iheyensis</i>	e-value
<i>Bacillus halodurans</i>	e-115
<i>Bacillus stearothermophilus</i>	2e-91
<i>Staphylococcus aureus</i> Mu50	5e-72
<i>Exiguobacterium</i> sp. 255	5e-72
<i>Staphylococcus aureus</i> MW2	7e-72
<i>Staphylococcus epidermidis</i>	2e-68
<i>Bacillus thuringiensis</i>	3e-38
<i>Bacillus anthracis</i> A2012	5e-38
<i>Bacillus anthracis</i> Ames	5e-38
<i>Bacillus cereus</i> ATCC 10987	1e-37
<i>Bacillus cereus</i> G9241	4e-37
<i>Bacillus cereus</i> ATCC 14579	2e-36
<i>Bacillus cereus</i> G9241	4e-36
<i>Bacillus halodurans</i>	8e-34
<i>Bacillus subtilis</i>	1e-33
<i>Bacillus cereus</i> ATCC 14579	2e-16
<i>Bacillus thuringiensis</i>	2e-16
<i>Bacillus cereus</i> ATCC 10987	4e-16
<i>Bacillus cereus</i> G9241	9e-16

search, all the detectable hits are either in *Bacillus* strains or in *Staphylococcus* strains (Table 1). Another *Bacillus* species, *B. stearothermophilus*, has a subunit II gene that shares high similarity with the one in *O. iheyensis*. This gene has been reported to function as a cytochrome *bd* oxidase subunit II (Sakamoto et al. 1999). The cytochrome *bd* subunit sequences from *B. stearothermophilus* are not included here, because only fully sequenced genomes are included in this study to ensure that all potential candidates are found. Unlike *B. halodurans* and *B. stearothermophilus*, all other *Bacillus* species show a fairly distant relation with *O. iheyensis* subunit II gene. Here, the cytochrome *bd* genes closely related to those of *O. iheyensis* are labeled the *Staphylococcus-Bacillus* cluster (as shown in Figs. 3 and 4). It is worth clarifying that the *Staphylococcus-Bacillus* cluster does not necessarily contain all cytochrome *bd* genes from the *Staphylococcus* or *Bacillus* genera. Genes from some other *Bacillus* species are not in the *Staphylococcus-Bacillus* cluster. Both subunits in the *Staphylococcus-Bacillus* cluster are closely related to genes in *Streptomyces avermitilis* (Figs. 3 and 4). However, there is a long branch leading to the *Staphylococcus-Bacillus* cluster in the phylogeny of subunit II genes (Fig. 4).

The Q-loop in subunit I was checked in the aligned sequences. There are 43 subunit I sequences in this study having a Q-loop between the putative transmembrane segments VII and VIII. Similarly, the corresponding 43 subunit II sequences all have about 30 more amino acids at the N-terminal between the putative loops V and VII than the remaining sequences. The difference in the overall size of the two

subunits is mainly due to the length of these two regions. But small deletions occurred during subunit I evolution in the *Staphylococcus-Bacillus* cluster. The subunit I genes from the *Staphylococcus-Bacillus* cluster show a significantly shorter sequence length than those from other non-Q-loop bacteria (Table 2).

Among bacterial genomes, in total, there are 70 species that have both homologues with *E. coli* K12 *cydAB*. The best hit of the *cydA* gene or *cydB* gene in each species was extracted respectively. Because only adjacently linked pairs were used to conduct pairwise distance analyses, a total of 68 bacterial species were used in the pairwise distance analyses. The average pairwise distance of subunit I genes among the 68 bacterial species is 1.31 ± 0.03 , whereas the average pairwise distance of subunit II genes among the 68 species is 1.97 ± 0.07 . This indicates that subunit II has a higher evolutionary rate on some branches. In the broader survey including Archaea, if more than one complex is present in a genome, all potential adjacently linked subunits are analyzed. In total, there are 114 pairs of genes found in 78 genomes (76 bacteria and 2 archaea). In these 114 pairs of genes, the average pairwise distance of genes encoding subunit I is 1.44 ± 0.02 , whereas the average pairwise distance of genes encoding subunit II is 2.07 ± 0.05 . The cytochrome *bd* subunit genes in the six selected species were also analyzed separately. The average pairwise distance of subunit I genes among the six species is 0.69 ± 0.10 , while the average pairwise distance of subunit II genes among the six species is 0.79 ± 0.14 . Figure 2 shows that the two subunits have the same topology. The phylogenetic topology of all significant homologues between the two subunits were compared by ProML (Felsenstein 1989) using the subunit II dataset and no significant difference was found between the phylogenetic topologies of the two subunits (Figs. 3 and 4) was found. The phylogenies of the best hits in each species between the two subunits also show similar topology (data not shown). It is striking that the two subunits of cytochrome *bd* in each species show similar evolutionary histories. However, the different distances of the two subunits indicate an asymmetrical evolutionary pattern of cytochrome *bd* subunits. It is worth noting that *appCB* were also used as query sequences and the results were not qualitatively different.

The distance-matrix rate test (Syvanen 2002) shows that subunit I, in general, evolves 1.2 time faster than subunit II (Fig. 5). The pairs lying above the 95% expectation level are related to the members from either the *Staphylococcus-Bacillus* cluster or ϵ -proteobacteria. Nevertheless, the pairs between the members within the *Staphylococcus-Bacillus* cluster or within ϵ -proteobacteria lie within the 95% confidence interval. Tajima's relative rate test shows that the subunit I genes in the *Staphylococcus-Bacillus*

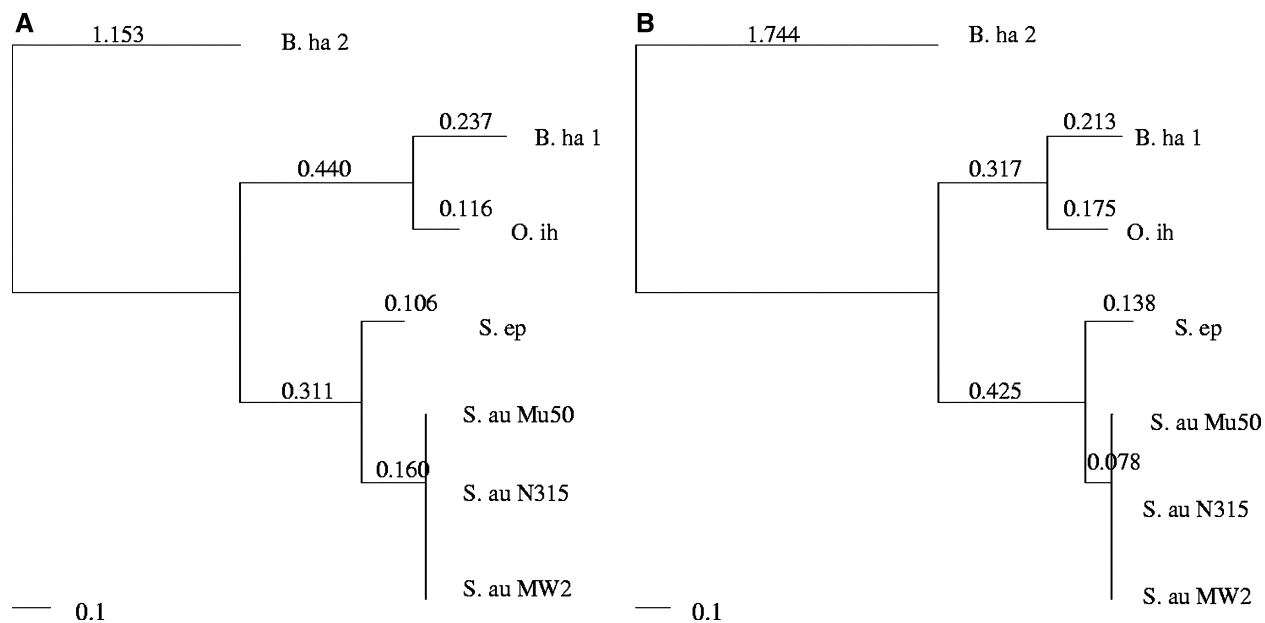


Fig. 2. Maximum likelihood unrooted tree of subunit I genes (**A**) and subunit II genes (**B**) from the six species (boxed in Fig. 1), that do not have homologues with *appB* or *cydB* of *E. coli* K12. The subunit II genes were extracted according to the KEGG and EcoCyc databases and confirmed by a BLASTP search. Branch

lengths are shown as the substitution rate per site. Both subunit genes in *S. aureus* Mu50 and *S. aureus* N315 are identical. All branches are supported by bootstrap values of at least 80% excluding those to the individual *S. aureus* isolates.

cluster evolve slightly slower than that on the *Streptomyces avermitilis* branch at the protein level (Table 3). However, the evolutionary rate of subunit II in the *Staphylococcus-Bacillus* cluster is greater than that on the *S. avermitilis* branch. For instance, the rate on the branch leading to the *Staphylococcus-Bacillus* group strain is from 2.18 times to 10.72 times greater than the rate on the *S. avermitilis* branch at the protein level. A similar result is found in the ϵ -proteobacteria group (Table A.1). The subunit II genes in ϵ -proteobacteria evolve over two times faster than those in non- ϵ -proteobacteria within the Q-loop cluster, while only a slight difference (1.22-fold difference) between ϵ -proteobacteria and non- ϵ -proteobacteria was detected in subunit I at the protein level. Similar results were found at the DNA level (Table 3). The accumulated effect of the elevated rate leading to the *Staphylococcus-Bacillus* cluster causes subunit II in the cluster to fall below significance levels in the sequence similarity search.

In the broader survey, the *E. coli cydAB* genes were used to BLAST against all of the complete archaeal genomes. Of the 19 complete archaeal genomes, two species have been detected to have genes encoding both subunits. They are *Methanosarcina acetivorans* and *Halobacterium* sp. NRC-1 (shown in Figs. 3 and 4). *M. acetivorans* has cytochrome *bd* genes, whereas the *cydAB* homologues were not found in the other *Methanosarcina* species: *M. mazei*. The cytochrome *bd* genes in *M. acetivorans* cluster with genes of *Bacillus cereus* and *B. anthracis*, suggesting that they possibly

have been laterally transferred from *Bacillus*. Four sets of cytochrome *bd* genes were found from the two plasmids in *Halobacterium* sp. NRC-1. They all have a close relation with genes in *Chlamydia*s and Firmicutes, but they are distantly related to the genes from *M. acetivorans*. This might be due to a LGT event from bacteria to archaea. Seven other subunit I homologues have been detected in four other archaeal genomes but there are no obvious subunit II homologues in them. These genomes are *Archaeoglobus fulgidus*, *Picrophilus torridus*, *Thermoplasma volcanium*, and *Thermoplasma acidophilum*. To further check for potential homologues, 14 genes, 1 on each side of the 7 subunit I homologues, were extracted and used as query sequences to BLAST against all detected subunit II genes. Seven genes show a weak match with known subunit II genes. Moreover, they show a shorter sequence length than bacterial genes (Table 2). This suggests that the subunit II genes in these four genomes are highly diverged from bacterial forms, which is consistent with our previous results. Again, the two subunits show an asymmetrical evolutionary pattern.

Discussion

The phylogenetic topology of the cytochrome *bd* complex shows that the two subunits have essentially similar topologies. However, pairwise distance analyses using gene pairs adjacent to each other show that

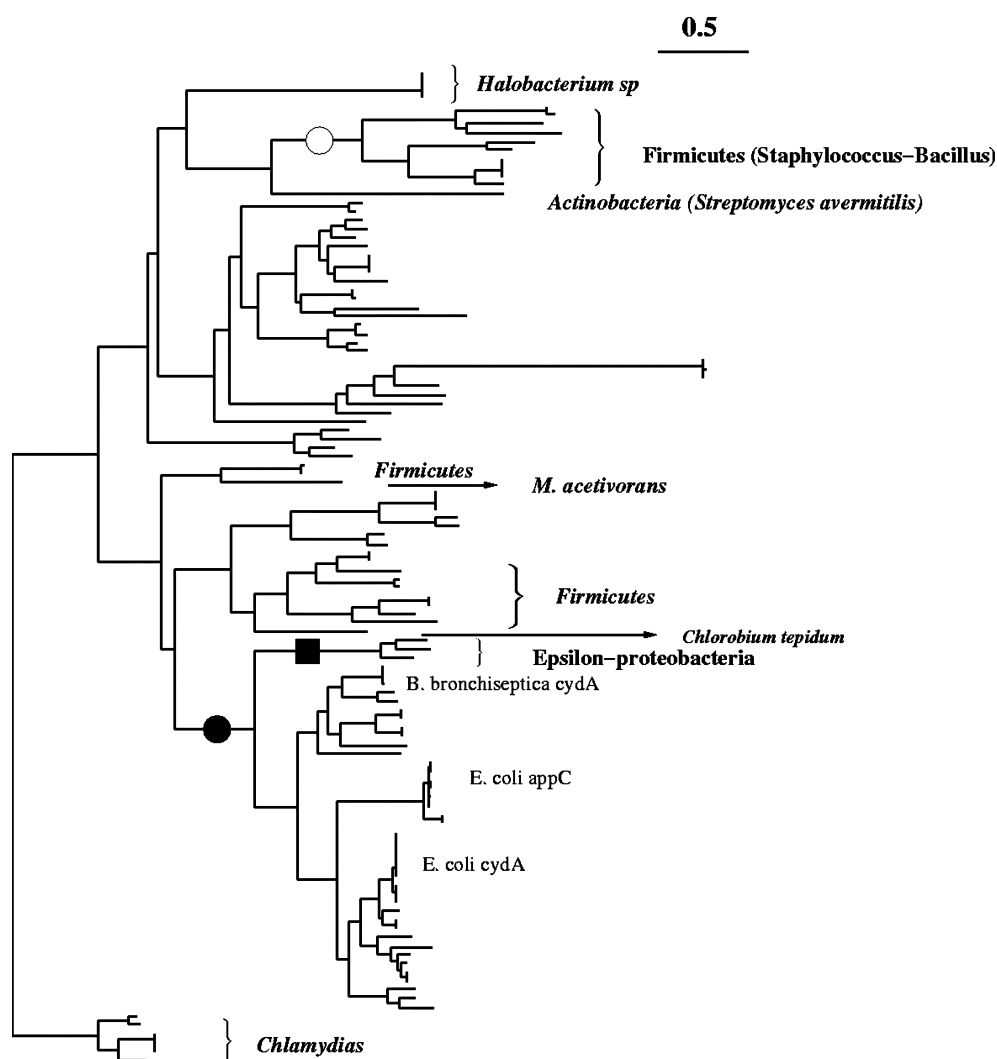


Fig. 3. Unrooted tree of cytochrome *bd* subunit I genes in a broad survey using all the significant hits to *E. coli cydA* as query sequences. The open circle indicates the origin of the *Staphylococcus-Bacillus* cluster, the filled circle indicates the origin of the Q-loop cluster, and the filled box indicates the origin of ϵ -proteobacteria.

the subunit II genes have accelerated evolutionary changes relative to subunit I genes. The BLAST survey against archaeal genomes shows four genomes have seven subunit I homologues but no clear subunit II homologues. Probably subunit II genes in these distantly related genomes are too diverged and have become undetectable. Indeed, there are seven genes that are found to have a weak match with other subunit II genes. All of these seven genes are adjacent to the seven subunit I homologues. While this supports a conclusion of asymmetrical evolution of these subunits, there is no experimental evidence in the literature to demonstrate an active subunit II in archaeal species.

Phylogenetic analyses of cytochrome *bd* genes show that some species in the phylogeny fail to support the 16sRNA tree (the 16sRNA phylogeny is not shown in this study), which has been employed as one way to detect LGT of metabolic enzymes (Brown and

Doolittle 1997; Osborne and Gennis 1999). For example, some genes are found in this study to have been most probably transferred from bacteria to archaea. The cytochrome *bd* genes in archaeal species *M. acetivorans* are closely related to bacterial species *B. anthracis* and *B. cereus*. Similarly, the close relationship of *Staphylococcus*, *O. iheyensis*, and *B. halodurans* might indicate a LGT event between *Bacillus* and *Staphylococcus*. Taken together, it might suggest a high degree of lateral gene transfer of cytochrome *bd* related to *Bacillus*. Cytochrome *bd* genes in *Halobacterium* sp. might also undergo LGT events from bacteria to archaea.

Besides these two archaeal species, four other archaeal species have genes encoding proteins with detectable similarity to subunit I and weak similarity to known subunit II genes. The remaining 13 archaeal species (strains) in this study are missing cytochrome *bd* subunits. The cytochrome *bd* subunits in

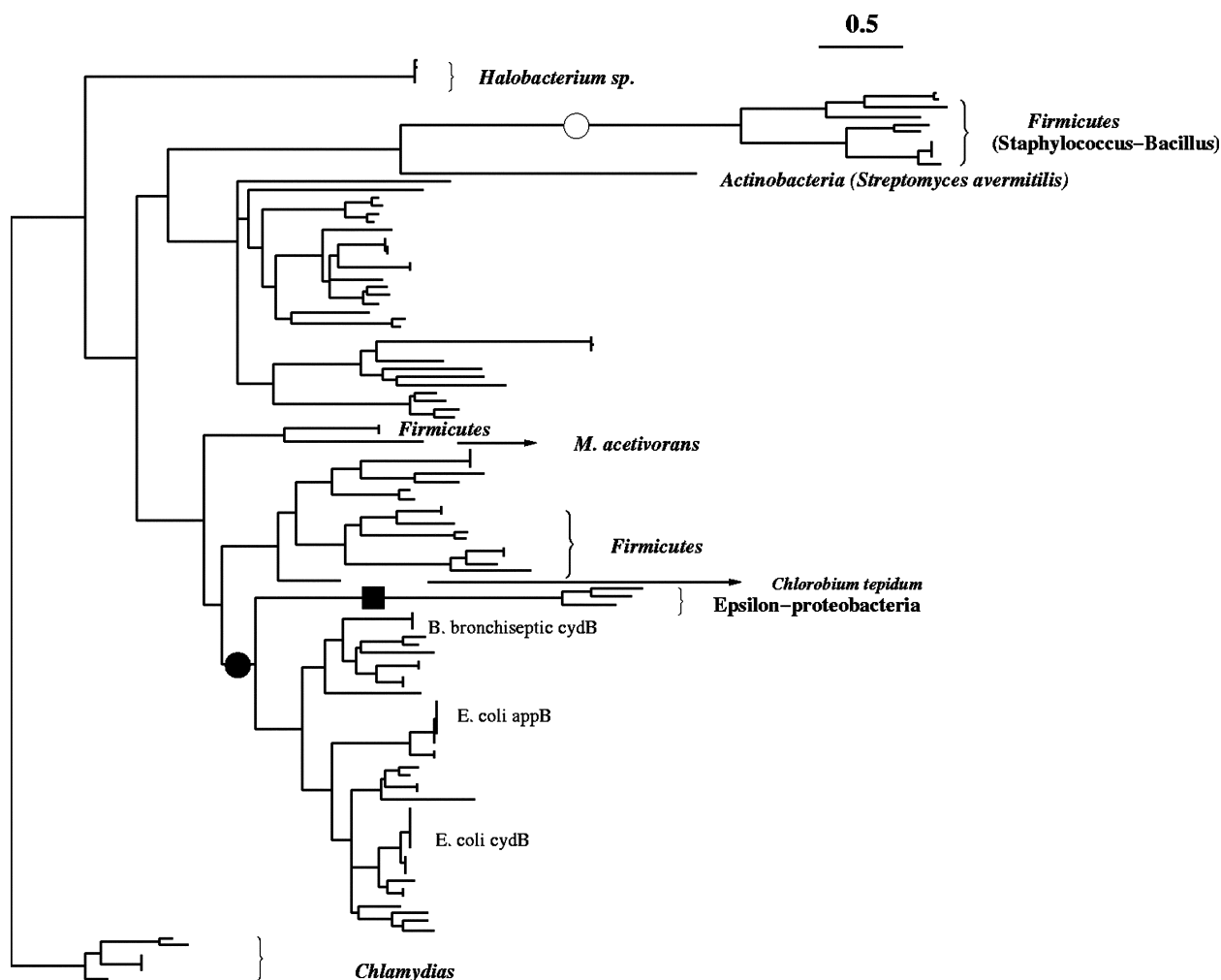


Fig. 4. Unrooted tree of cytochrome *bd* subunit II genes in a broad survey using all the significant hits to *E. coli cydB* as query sequences. The open circle indicates the origin of the *Staphylococcus-Bacillus* cluster, the filled circle indicates the origin of the Q-loop cluster, and the filled box indicates the origin of ϵ -proteobacteria. The evolutionary history is shown on the same scale in Fig. 3.

Table 2. The average length in amino acids (along with their standard errors) of cytochrome *bd* genes among Archaea, the *Staphylococcus-Bacillus* group, the Q-loop group, and all other bacteria

	Archaea ^a	SB group ^b	Q-loop ^c	Bacteria ^d
Subunit I	473.2 \pm 7.1	448.8 \pm 1.6	524.5 \pm 1.4	466.1 \pm 3.2
Subunit II	331.0 \pm 3.9	340.5 \pm 1.0	378.7 \pm 0.7	339.4 \pm 0.8

^aIn addition to *Halobacterium* sp. and *M. acetivorans* (shown in Figs. 3 and 4), the genes from *A. fulgidus*, *P. torridus*, *T. volcanium*, and *T. acidophilum* are included here, despite the weak match of their subunit II genes to genes from other species.

^bThis group contains genes in the *Staphylococcus-Bacillus* cluster (shown in Figs. 3 and 4).

^cThis group contains all cytochrome *bd* complexes which have Q-loop in subunit I (shown in Figs. 3 and 4).

^dThis group contains all the bacterial genes excluding the ones from the SB group and the Q-loop group.

these four archaeal species may have a different origin from the bacterial genes. Cytochrome *bd* subunits are absent in many archaeal species.

It is worth noting that the results were not qualitatively different using *cydAB* and *appCB*. *cydAB* genes and *appCB* genes are present sporadically on the phylongey (data not shown). The name (e.g. *app* or *cyd*) of cytochrome *bd* subunits, therefore, does not necessarily reflect their phylogenetic relationship.

Surprisingly, the similarity of subunit II (either *cydB* or *appB*) between *E. coli* K12 versus *O. iheyensis*, *B. halodurans*, and other four *Staphylococcus* species at the protein level is almost undetectable. And yet the subunit II genes in these six species have been either annotated or experimentally confirmed to function as cytochrome *bd* subunit II (Sakamoto et al. 1999; Takami et al. 2000; Lu et al. 2001; Kuroda et al. 2001).

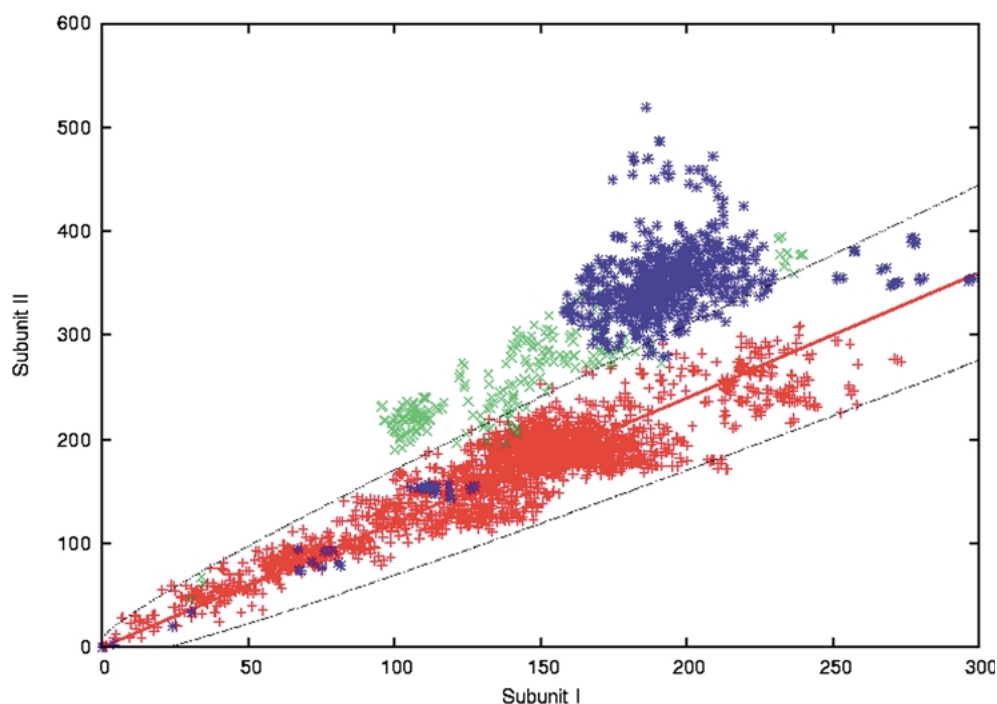


Fig. 5. Distance-matrix rate test plot between subunit I and subunit II. The straight line is $d_y = 1.2d_x$ and the dotted lines outline the 95% confidence levels. The numbers d_y and d_x are the number of replacements per 100 residues estimated from the se-

quences. The pairs having members of the *Staphylococcus-Bacillus* cluster are shown in blue; pairs having either *Campylobacter jejuni*, *Helicobacter hepaticus*, or *Wolinella succinogenes*, in green; and pairs within the remaining bacteria, in red.

The distance-matrix rate test (Syvanen 2002) was used to test the relative rates of subunit I and subunit II. The distance-matrix rate test makes no assumptions regarding divergence times, or phylogenetic relationships. In fact, the subunit phylogenies do not always support the species phylogenies due to the frequency of the LGT events of the whole operon. Tajima's (1993) relative rate test was used to test specific lineage elevated evolutionary rates, since the relative rate test is more powerful than the likelihood ratio test when one set of descendants has evolved faster than the other set (Sorhannus and Van Bell 1999). The power of relative rate test is dependent on the number of variable sites considered, the rate of substitution, and the distance from the ingroup to the outgroup (Tajima 1993; Bromham et al. 2000). The tests show that subunit II genes in the *Staphylococcus-Bacillus* cluster and ϵ -proteobacteria evolve significantly faster than their neighbors. However, subunit I genes in the *Staphylococcus-Bacillus* cluster evolve slower than their neighbor *S. avermitilis* and subunit I genes in ϵ -proteobacteria only evolve slightly faster than their neighbors. For Tajima's relative rate test, the outgroup of the *Staphylococcus-Bacillus* cluster versus *S. avermitilis* and ϵ -proteobacteria versus non- ϵ -proteobacteria in the Q-loop cluster was chosen based on the relatedness on the phylogenies (Figs. 3 and 4). Several outgroups were tested and the results did not significantly differ (data not shown). Three

different operons from non- ϵ -proteobacteria were compared with ϵ -proteobacteria and the trend of elevated evolution held true. Hence, the robustness of the elevated evolution of subunit II on the branches leading to the *Staphylococcus-Bacillus* cluster and ϵ -proteobacteria is largely independent of the genes sampled.

To understand any possible change of selection pressure, the PAML package (Yang 1997) was used to conduct a comparison of K_a/K_s (ω value) between each group and the outgroup. No significant difference among subunit II genes has been detected on the two groups which show elevated evolution in likelihood ratio test (data not shown). Different ω values were detected on the branch leading to the *Staphylococcus-Bacillus* cluster of subunit I (data not shown), but importantly the subunit I genes in this cluster show a significantly lower ratio to the outgroup branch in the relative rate tests using protein and DNA sequences (Table 3). This suggests that the elevated evolution of subunit II genes in the *Staphylococcus-Bacillus* cluster and the ϵ -proteobacteria group could not be simply explained by a change of selection pressure.

The BLAST search results show that the novel subunit II gene in *Oceanobacillus iheyensis* only has significant hits with *Bacillus* and *Staphylococcus* (Table 1). The novel subunit II gene has homologues in *B. subtilis*, *B. cereus*, and *B. anthracis*. These homologues have detectable hits in *Streptomyces*

Table 3. Relative rate test between the *Staphylococcus-Bacillus* cluster and *Streptomyces avermitilis*: One copy of relative subunit from *Halobacterium* sp. NRC-1 was used as the outgroup in the relative rate test of each subunit

Taxon	Subunit I				Subunit II			
	Protein		DNA		Protein		DNA	
	m ₁ /m ₂	χ ²	m ₁ /m ₂	χ ²	m ₁ /m ₂	χ ²	m ₁ /m ₂	χ ²
1	0.59	52.88***	0.98	0.04	2.18	208.34***	1.80	29.46***
2	0.61	43.81***	1.11	1.11	2.51	269.09***	2.91	88.37***
3	0.65	34.14***	1.33	7.51**	2.96	369.26***	6.56	238.26***
4	0.66	34.14***	1.33	7.51**	2.96	369.26***	6.56	238.26***
5	0.66	33.12***	1.33	7.51**	2.95	367.91***	5.71	202.24***
6	0.68	29.34***	1.33	7.47**	2.97	387.57***	3.58	111.79***
7	1.05	0.60	1.53	16.99***	6.73	718.86***	3.66	134.80***
8	1.05	0.50	1.61	21.27***	5.79	650.15***	2.90	89.70***
9	0.85	4.89*	1.16	2.33	10.72	937.24***	2.85	98.24***
10	0.91	1.81	1.21	3.59	2.84	312.70***	1.94	38.74***
Average	0.77		1.29		4.26		3.85	

Note. Significant at * 5% level, ** 1% level, and *** 0.1% level (1 df).

avermitilis, which is also the near neighbor of *Bacillus* according to the subunit I phylogeny (Figs. 3 and 4). The novel subunit II genes in each species are always adjacent to subunit I genes on chromosome. This suggests that the origin of the novel subunit II genes in *Bacillus* and *Staphylococcus* is due to vertical evolution and accelerated evolution of subunit II genes. However, the possibility of LGT of subunit II genes cannot be completely ruled out. In fact, LGT of individual subunits has been reported in a multiple-component complex (Iyer et al. 2004).

It is apparent that other *Bacillus* species such as *B. cereus*, *B. anthracis*, and *B. subtilis* have more than one copy of cytochrome *bd* genes. During evolution, the lineage to *O. iheyensis* and *B. halodurans* may have lost the cytochrome *bd* genes which have greater similarity with *E. coli cydAB*. Duplicate genes often have rapid divergence after duplication (Gu et al. 2002) and diversification after duplication is often asymmetrical (Wagner 2002; Conant and Wagner 2003; Zhang et al. 2003). Another possible explanation is that the subunit II genes in the six species were analogously displaced by unrelated sequences (non-orthologous replacement). Indeed, displacement of an enzyme is now widely acknowledged as an important source of phylogenetic difference (Dhillon et al. 2003; Morett et al. 2003). All subunit II genes from the *Staphylococcus-Bacillus* cluster were used as query sequences to find potential homologues without regard to the physical location relative to subunit I. There were no new homologues found as near neighbors of the *Staphylococcus-Bacillus* cluster. This suggests that the observed novel subunit II genes are not an artifact of the methodology which only considers physically adjacent subunits. The novel subunit II genes in the *Staphylococcus-Bacillus* cluster are more likely due to elevated evolution after duplication and the original gene cluster is missing, but the

possibility of nonorthologous replacement cannot be ruled out.

Similar studies have been done by Osborne and Gennis (1999) and Sakamoto et al. (1999). Osborne and Gennis (1999) generated phylogenetic trees using sequences without the Q-loop region. They showed that genes with the Q-loop cluster within one clade of the tree. Our study using the whole sequence alignment shows a similar result (Figs. 3 and 4). It shows that all the 43 sequences with the Q-loop cluster entirely in one clade of the tree, and that all of them are from proteobacteria. Sakamoto et al. (1999) constructed their phylogenetic trees using only several selected fragments of highly conserved regions. Their work showed an essentially similar rate of evolution between the two subunits. The subunit II gene from *B. stearothermophilus*, which has high similarity to subunit II gene in *O. iheyensis* (Table 1), does not show an accelerated evolutionary rate in their work. This is due to the limited distance information provided by the small selected region chosen. Our study using the whole sequence alignment and a broader data set shows considerable asymmetrical evolution of the two cytochrome *bd* subunits.

Rate heterogeneity and rate variation have been reported among cytochrome *c* oxidase subunits (Spicer 1995; Schmidt et al. 1997, 2002). In these studies, subunits were compared within closely related eukaryotic species. Lineage-specific elevated rates of evolution of cytochrome *c* oxidase in primates have been reported (Grossman et al. 2001). The subunits have been shown to have co-occurring elevated rates in primates (Andrews et al. 1998; Andrews and Easteal 2000; Wu et al. 2000; Schmidt et al. 2001; Goldberg et al. 2003). In other words, the subunit genes that are known to interact functionally show a “coevolution” pattern. Our study shows an interesting case in related subunits where two su-

bunits evolve differently and that dramatically elevated evolution might cause homologues to become undetectable in a sequence similarity search. The results also imply that a sequence similarity search might fail completely and provide misleading results due to lineage specific elevated and asymmetrical evolution.

The asymmetrical evolution of subunits in the two-component complex cytochrome *bd* observed here is common on most branches. A relaxation of selective constraints may be responsible for the asymmetrical evolution of subunits. Indeed, several conserved residues in subunit I have been suggested to be involved in binding heme (Osborne and Gennis 1999; Sakamoto et al. 1999). The way in which subunit II interacts with heme is still unclear. To date, there is no available crystal structure data of cytochrome *bd*. When the crystal structure data become available, the different constraints between the two subunits will become clearer.

Acknowledgment. This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) grant to G.B.G.

References

- Altschul SF, Madden TL, Schaffer 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
- Andrews TD, Eastal S (2000) Evolutionary rate acceleration of cytochrome *c* oxidase subunit I in simian primates. *J Mol Evol* 50:562–568
- Andrews TD, Jermin LS, Eastal S (1998) Accelerated evolution of cytochrome *b* in simian primates: adaptive evolution in concert with other mitochondrial proteins? *J Mol Evol* 47:249–257
- Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–1474
- Bromham L, Penny D, Rambaut A, Hendy MD (2000) The power of relative rates tests depends on the data. *J Mol Evol* 50:296–301
- Brown JR, Doolittle WF (1997) Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev* 61:456–502
- Ciccarese S, Lanave C, Saccone C (1997) Evolution of T-cell receptor gamma and delta constant region and other T-cell-related proteins in the human-rodent-artiodactyl triplet. *Genetics* 145:409–419
- Conant GC, Wagner A (2003) Asymmetric sequence divergence of duplicate genes. *Genome Res* 13:2052–2058
- Dassa J, Fsihi H, Marck C, Dion M, Kieffer-Bontemps M, Boquet PL (1991) A new oxygen-regulated operon in *Escherichia coli* comprises the genes for a putative third cytochrome oxidase and for pH 25 acid phosphatase (*appA*). *Mol Gen Genet* 229:341–352
- Dhillon IS, Marcotte EM, Roshan U (2003) Diametrical clustering for identifying anti-correlated gene clusters. *Bioinformatics* 19:1612–1619
- Felsenstein J (1989) PHYLIP (phylogeny inference package). *Version 32 Cladistics* 5:164–166
- Ghaim JB, Greiner DP, Meares CF, Gennis RB (1995) Proximity mapping the surface of a membrane protein using an artificial protease: demonstration that the quinone-binding domain of subunit I is near the N-terminal region of subunit II of cytochrome *bd*. *Biochemistry* 34:11311–11315
- Gogarten JP, Doolittle WF, Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* 19:2226–2238
- Goldberg A, Wildman DE, Schmidt TR, Huttemann M, Goodman M, Weiss ML, Grossman LI (2003) Adaptive evolution of cytochrome *c* oxidase subunit VIII in anthropoid primates. *Proc Natl Acad Sci USA* 100:5873–5878
- Green GN, Kranz RG, Lorence RM, Gennis RB (1984) Identification of subunit I as the cytochrome *b₅₅₈* component of the cytochrome *d* terminal oxidase complex of *Escherichia coli*. *J Biol Chem* 259:7994–7997
- Green GN, Lorence RM, Gennis RB (1986) Specific overproduction and purification of the cytochrome *b₅₅₈* component of the cytochrome *d* complex from *Escherichia coli*. *Biochemistry* 25:2309–2314
- Grossman LI, Schmidt TR, Wildman DE, Goodman M (2001) Molecular evolution of aerobic energy metabolism in primates. *Mol Phylogenet Evol* 18:26–36
- Gu Z, Nicolae D, Lu HH, Li WH (2002) Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet* 18:609–613
- Gutierrez RA, Larson MD, Wilkerson C (2004) The plant-specific database. Classification of Arabidopsis proteins based on their phylogenetic profile. *Plant Physiol* 135:1888–1892
- Iyer LM, Koonin EV, Aravind L (2004) Evolution of bacterial RNA polymerase: implications for large-scale bacterial phylogeny, domain accretion, and horizontal gene transfer. *Gene* 335:73–88
- Jain R, Rivera MC, Lake JA (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci USA* 96:3801–3806
- Jordan IK, Marino-Ramirez L, Wolf YI, Koonin EV (2004) Conservation and co-evolution in the scale-free human gene coexpression network. *Mol Biol Evol* 21:2058–2070
- Junemann S (1997) Cytochrome *bd* terminal oxidase. *Biochim Biophys Acta* 1321:107–127
- Karp PD, Riley M, Saier M, Paulsen IT, Collado-Vides J, Paley SM, Pellegrini-Toole A, Bonavides C, Gama-Castro S (2002) The EcoCyc Database. *Nucleic Acids Res* 30:56–58
- Koski LB, Golding GB (2001) The closest BLAST hit is often not the nearest neighbor. *J Mol Evol* 52:540–542
- Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki K, Nagai Y, Lian J, Ito T, Kanamori M, Matsumaru H, Maruyama A, Murakami H, Hosoyama A, Mizutani-Ui Y, Takahashi NK, Sawano T, Inoue R, Kaito C, Sekimizu K, Hirakawa H, Kuhara S, Goto S, Yabuzaki J, Kanehisa M, Yamashita A, Oshima K, Furuya K, Yoshino C, Shiba T, Hattori M, Ogasawara N, Hayashi H, Hiramatsu K (2001) Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* 357:1225–1240
- Lawrence JG, Ochman H (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci USA* 95:9413–9417
- Lu J, Nogi Y, Takami H (2001) *Oceanobacillus iheyensis* gen. nov., sp. nov., a deep-sea extremely halotolerant and alkaliphilic species isolated from a depth of 1050 m on the Iheya Ridge. *FEMS Microbiol Lett* 205:291–297
- Marcotte EM, Xenarios I, van Der Blik AM, Eisenberg D (2000) Localizing proteins in the cell from their phylogenetic profiles. *Proc Natl Acad Sci USA* 97:12115–12120
- Mikkelsen TS, Galagan JE, Mesirov JP (2005) Improving genome annotations using phylogenetic profile anomaly detection. *Bioinformatics* 21:464–470
- Morett E, Korb J, Rajan E, Saab-Rincon G, Olvera L, Olvera M, Schmidt S, Snel B, Bork P (2003) Systematic discovery of

- analogous enzymes in thiamin biosynthesis. *Nat Biotechnol* 21:790–795
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 27:29–34
- Osborne JP, Gennis RB (1999) Sequence analysis of cytochrome *bd* oxidase suggests a revised topology for subunit I. *Biochim Biophys Acta* 1410:32–50
- Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci USA* 96:4285–4288
- Sakamoto J, Koga E, Mizuta T, Sato C, Noguchi S, Sone N (1999) Gene structure and quinol oxidase activity of a cytochrome *bd*-type oxidase from *Bacillus stearothermophilus*. *Biochim Biophys Acta* 1411:147–158
- Schmidt TR, Jaradat SA, Goodman M, Lomax MI, Grossman LI (1997) Molecular evolution of cytochrome *c* oxidase: rate variation among subunit VIa isoforms. *Mol Biol Evol* 14:595–601
- Schmidt TR, Wu W, Goodman M, Grossman LI (2001) Evolution of nuclear- and mitochondrial-encoded subunit interaction in cytochrome *c* oxidase. *Mol Biol Evol* 18:563–569
- Schmidt TR, Goodman M, Grossman LI (2002) Amino acid replacement is rapid in primates for the mature polypeptides of COX subunits, but not for their targeting presequences. *Gene* 286:13–19
- Sorhannus U, Van Bell C (1999) Testing for equality of molecular evolutionary rates: a comparison between a relative-rate test and a likelihood ratio test. *Mol Biol Evol* 16:849–855
- Spicer GS (1995) Phylogenetic utility of the mitochondrial cytochrome oxidase gene: molecular evolution of the *Drosophila buzzatii* species complex. *J Mol Evol* 41:749–759
- Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum-likelihood for reconstructing tree topologies. *Mol Biol Evol* 13:964–969
- Sturr MG, Krulwich TA, Hicks DB (1996) Purification of a cytochrome *bd* terminal oxidase encoded by the *Escherichia coli* *app* locus from a delta *cyo* delta *cyd* strain complemented by genes from *Bacillus firmus* OF4. *J Bacteriol* 178:1742–1749
- Syvanen M (2002) Rates of ribosomal RNA evolution are uniquely accelerated in eukaryotes. *J Mol Evol* 55:85–91
- Tajima F (1993) Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135:599–607
- Takami H, Nakasone K, Takaki Y, Maeno G, Sasaki R, Masui N, Fuji F, Hiramata C, Nakamura Y, Ogasawara N, Kuhara S, Horikoshi K (2000) Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res* 28:4317–4331
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Timkovich R, Cork MS, Gennis RB, Johnson PY (1985) Proposed structure of heme *d*, a prosthetic group of bacterial terminal oxidases. *J Am Chem Soc* 107:6069–6075
- Trumpower BL, Gennis RB (1994) Energy transduction by cytochrome complexes in mitochondrial and bacterial respiration: the enzymology of coupling electron transfer reactions to transmembrane proton translocation. *Annu Rev Biochem* 63:675–716
- Wagner A (2002) Asymmetric functional divergence of duplicate genes in yeast. *Mol Biol Evol* 19:1760–1768
- Wu W, Schmidt TR, Goodman M, Grossman LI (2000) Molecular evolution of cytochrome *c* oxidase subunit I in primates: Is there coevolution between mitochondrial and nuclear genomes? *Mol Phylogenet Evol* 17:294–304
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556
- Zhang P, Gu Z, Li WH (2003) Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biol* 4:R56
- Zheng Y, Roberts RJ, Kasif S (2002) Genomic functional annotation using co-evolution profiles of gene clusters. *Genome Biol* 3:research0060