Molecular Phylogenetics and Evolution 69 (2013) 674–686

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Elucidating the origin of the ExbBD components of the TonB system through Bayesian inference and maximum-likelihood phylogenies

Livingstone Marmon *

1326 Mountain Vista Dr SE, Rio Rancho, New Mexico 87124, United States

ARTICLE INFO

Article history:
Received 25 July 2012
Revised 28 June 2013
Accepted 12 July 2013
Available online 23 July 2013

Keywords:
TonB system
Molecular phylogenetics
Flagellar motor
Homology
Molecular evolution

ABSTRACT

Uptake of ferric siderophores, vitamin B12, and other molecules in gram-negative bacteria is mediated by a multi-protein complex known as the TonB system. The ExbB and ExbD protein components of the TonB system play key energizing roles and are homologous with the flagellar motor proteins MotA and MotB. Here, the phylogenetic relationships of ExbBD and MotAB were investigated using Bayesian inference and the maximum-likelihood method. Phylogenetic trees of these proteins suggest that they are separated into distinct monophyletic groups and have originated from a common ancestral system. Several horizontal gene transfer events for ExbB–ExbD are also inferred, and a model for the evolution of the TonB system is proposed.

1. Introduction

Active transport of iron ions in gram-negative bacteria is catalyzed by a multi-protein complex known as the TonB system (Braun, 1995). Located in the cytoplasmic membrane (Braun, 1995), this system is composed of three inner membrane protein components: TonB, ExbB, and ExbD (Braun et al., 1996; Krewulak and Vogel, 2011; Noinaj et al., 2010). Although it was initially thought that TonB-dependent transport was limited to iron complexes and vitamin B12 (Schauer et al., 2008), bioinformatics and other approaches have demonstrated that ligands of TonB-dependent receptors include sugars, heme, and non-ferrous cations (Lim, 2010).

The TonB system energizes active transport in the cell by way of the proton-motive force (PMF) (Jana et al., 2011; Ollis and Postle, 2011; Ollis et al., 2009; Postle and Kadner, 2003; Swayne and Postle, 2011). As a result of a number of experimental studies, the overall topologies and functions of the three TonB proteins have been elucidated. Three transmembrane (TM) domains are present in ExbB (Kampfenkel and Braun, 1993), while ExbD and TonB each have one TM domain (Hannavy et al., 1990; Kampfenkel and Braun, 1992; Ollis et al., 2009; Roof et al., 1991). ExbB and ExbD play key roles in transducing the PMF to TonB (Ollis et al., 2009), which undergoes conformational changes, transmitting potential energy to transporters in the outer membrane (Ghosh and Postle, 2005). Recent evidence, however, conflicts with this model of TonB function, and a new model of TonB function has been suggested (Gresock et al., 2011). For a review of the proposed mechanisms of energy transduction in TonB, see Krewulak and Vogel (2011).

Homology between the ExbB and ExbD components of the TonB system and the flagellar motor proteins MotA and MotB, respectively, has been previously noted (Cascales et al., 2001; Kojima and Blair, 2001; Pallen and Matzke, 2006; Zhai et al., 2003), and statistically significant sequence similarity between ExbB and MotA has been found (Kojima and Blair, 2001). Functional considerations also provide evidence of homology between the TonB complex and the flagellar motor. For instance, amino acid residues critical to ExbD function have been identified (see, e.g., Jana et al., 2011), including an aspartate residue at position 25 (Braun et al., 1996); likewise, the corresponding amino acid residue in MotB (D32) is important for the latter's activity (Ollis et al., 2009). Furthermore, the MotAB and the ExbBD complexes have similar topologies (Zhai et al., 2003).

What, then, is the nature of the phylogenetic relationship between the TonB complex and the flagellar system? In the following research, this question is examined from the angle of molecular phylogenetics. Maximum-likelihood (ML) and Bayesian inference are used to construct phylogenies of ExbB/MotA and ExbD/MotB, and the roots are inferred through the midpoint method and molecular clock analyses. The evolutionary history of ExbBD is also explored by comparing the phylogeny of these sequences to a species tree of bacteria phyla.

* Fax: +1 505 458 1512.
E-mail address: marmon.livingstone@gmail.com

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.07.010

2. Materials and methods

2.1. Protein sequence retrieval

Retrieval of ExbB/MotA and ExbD/MotB protein sequences was accomplished through the use of the NCBI protein sequence database (http://www.ncbi.nlm.nih.gov/protein). The bacteria phyla presented in Table 1 were searched for these protein sequences, using specific keywords. For example, to find ExbB sequences in Acidobacteria, the phrase "ExbB Acidobacteria" was employed. An identical strategy was used in the case of ExbD, MotA, and MotB. Since these searches often yielded ambiguous results (e.g., a protein might be listed as "ExbB_MotA"), particular annotation criteria were used to determine if a protein was to be classified as ExbB, ExbD, MotA, or MotB (see Table 2). A total of 32 ExbB/MotA and 29 ExbD/MotB sequences were collected through this procedure.

The accession numbers of the various sequences may be found in Appendix A.

2.2. Phylogenetic analyses

2.2.1. Multiple sequence alignment

Multiple sequence alignment (MSA) of ExbB/MotA and ExbD/MotB was executed through the MUSCLE program (Edgar, 2004) under default conditions. All columns containing gaps were removed from the MSAs. The best-fit models of protein evolution for the alignments were ascertained with ProtTest 2 (Abascal et al., 2005). For ExbB/MotA, the model with the highest overall ranking was the LG model (Le and Gascuel, 2008) with a proportion of invariable sites, a gamma distribution, and the empirical method for estimating amino acid frequencies (LG + I + G + F). In the case of ExbD/MotB, the LG + G model of evolution was ranked as the best overall.

For the construction of a phylogeny of ExbBD, ExbB and ExbD sequences were concatenated (e.g., the ExbD from Acidobacteria was joined with the ExbB from Acidobacteria) and aligned using the methodology described above. The best-fit model for this MSA, as determined by ProtTest 2 (Abascal et al., 2005), was LG + I + F.

2.2.2. Phylogenetic inference

A variety of methods are used in the reconstruction of phylogenetic trees, reviewed elsewhere (Baxevanis and Ouellette, 2001; Blair and Murphy, 2011; Clote and Backofen, 2000; Durbin et al., 1998; Felsenstein, 1996; Graur and Li, 2000; Jin et al., 2007; Maddison and Maddison, 2000; Moreira and Philippe, 2000; Nei, 1996; Thornton and DeSalle, 2000; Yang, 1996). Along with Bayesian inference, maximum-likelihood (ML) is among the most accurate tree-building methods (Philippe et al., 2011). Thus, this was one of the methods of choice for the construction of phylogenetic trees of ExbB/MotA and ExbD/MotB. The PhyML 3.0 (Guindon et al., 2010) interface provided by the HIV database (http://www.hiv.lanl.gov/content/sequence/PHYML/interface.html) was utilized for estimating the phylogenies of ExbB/MotA , ExbD/MotB, and ExbBD. The following parameters were used in all cases: firstly, the number of substitution rate categories was 4, the tree topologies and branch lengths of the starting trees (BioNJ) were optimized, tree improvement was effected through subtree pruning and regrafting, and the number of bootstrap replicates was set to 100.

The resulting ML phylogenies were rooted through the midpoint method as implemented in FigTree (http://tree.bio.ed.ac.uk/software/figtree/). This rooting method is often used when a clear outgroup is lacking (Boykin et al., 2010), and places the root of the tree at the midpoint of the two most divergent sequences (Hess and Russo, 2007).

A second rooting approach was also adopted for the ExbB/MotA and ExbD/MotB phylogenies: namely, the inference of the trees under the assumption of a molecular clock. Since the molecular clock model implies that the length from the root to all operational taxonomic units (OTUs) is the same, a phylogeny generated from a molecular clock perspective is rooted by definition (Boykin et al., 2010). Phylogeny estimation under the clock assumption was done in a Bayesian framework through the MrBayes program (version 3.2.1) (Ronquist et al., 2012). For both ExbB/MotA and ExbD/MotB, 5 million generations were initially run (25.0% burn-in) to test the strict clock model versus the non-clock model. It was thereby observed that the strict clock model was superior to the non-clock model in explaining the evolution of the datasets (see Table 3).

Under the strict clock model, Bayesian trees were estimated for each dataset. Given that MrBayes does not provide the option of the LG model of evolution, the WAG model was used instead (the WAG model is related to the LG model; see, e.g., Le and Gascuel, 2008). The MrBayes parameters for the ExbB/MotA alignment were as follows: prset aamodelpr = fixed(wag); lset rates = invgamma; prset brlenspr = clock:uniform; mcmc ngen = 5,000,000.

With regards to the ExbD/MotB alignment, these command lines were used: prset aamodelpr = fixed(wag); prset brlenspr = clock:uniform; lset rates = gamma; mcmc ngen = 5,000,000.

In total, then, five trees were constructed. The three ML trees (ExbB/MotA, ExbD/MotB, ExbBD) were rooted with the midpoint method, while the two Bayesian-inference phylogenies (ExbB/MotA, ExbD/MotB) were rooted through clock analyses.

2.3. Investigating the evolutionary history of ExbB/ExbD

Several approaches were used to further investigate the evolutionary history of ExbBD. To identify indications of horizontal gene transfer, the GC-content of the genes encoding the ExbB and ExbD sequences used in this study were calculated and compared to the GC- content of the genomes in which the genes were located. Also, phylogeny reconciliation of ExbBD with a species tree of bacteria phyla was done under two frameworks: horizontal gene transfer and the insertion of gene duplication and loss events.

2.3.1. Determining GC-content of genes and genomes

GC-content of the genes encoding the ExbB and ExBD sequences was determined by "EMBOSS 6.3.1: geecee" (http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::geecee) at Mobyle@Pasteur. The GC-content of the genomes from which each gene was found was estimated in the same way, unless the NCBI genome database (http://www.ncbi.nlm.nih.gov/genome) already listed the GC-content.

Table 1
Bacteria phyla that were searched for ExbBD/MotAB protein sequences using the NCBI protein sequence database.

Acidobacteria	Fibrobacteres
Actinobacteria	Firmicutes
Aquificae	Fusobacteria
Bacteroidetes	Gemmatimonadetes
Chlamydiae	Nitrospirae
Chlorobi	Planctomycetes
Chloroflexi	Proteobacteria
Chrysiogenetes	Spirochaetes
Cyanobacteria	Synergistetes
Deferribacteres	Tenericutes
Deinococcus-Thermus	Thermodesulfobacteria
Dictyoglomi	Thermotogae Verrucomicrobia

Table 2
The annotation criteria for selecting ExbBD/MotAB proteins in bacteria phyla are shown in this table.

Protein name	Annotation keywords
ExbB	Outer membrane transport energization protein ExbB, biopolymer transport protein exbB, ExbB, biopolymer transport ExbB protein, tonB-system energizer ExbB, TonB family auxiliary protein ExbB
ExbD	Outer membrane transport energization protein ExbD, biopolymer transport protein exbD, ExbD, exbD gene product, biopolymer transport exbD protein, TonB family auxiliary protein ExbD, ligand gated channel protein ExbD
MotA	Chemotaxis MotA protein, chemotaxis protein MotA, motility protein A, motA gene product, chemotaxis protein PomA, flagellar motor protein MotA, MotA, flagellar motor component, flagellar motor stator protein MotA
MotB	Chemotaxis protein MotB, motB gene product, flagellar motor protein MotB, MotB, chemotaxis MotB protein, flagellar motor protein

Table 3
The harmonic means are always lower for the strict clock model than for the non-clock model, suggesting that the strict clock model is better. Typically, a difference of 5 log units is regarded as strong evidence for the superior model.

Protein names	Harmonic mean: strict clock model (log units)	Harmonic mean: non-clock model (log units)
ExbB/MotA	−6308.69	−6313.21
ExbD/MotB	−5540.64	−5561.49

2.3.2. Species tree of bacteria phyla

Presently, there exists no clear consensus on the branching order of bacteria (Galtier and Daubin, 2008). Phylogenies of bacteria phyla vary depending on the molecular datasets that are employed (see, e.g., Gupta, 2001; Rappé and Giovannoni, 2003; Ciccarelli et al., 2006). For example, phylogenies based on 16S rRNA (e.g., Rappe and Giovanoni, 2003) are not always in agreement with phylogenies estimated from conserved proteins (e.g., Ciccarelli et al., 2006). The phylogeny of the latter study was chosen for use as the species tree (a phylogeny built from multiple conserved proteins seems likely to be more accurate than one generated from only one kind of molecule). Instead of reproducing the methods of Ciccarelli et al. (2006), a species tree of bacteria phyla was simply drawn after Ciccarelli et al. (2006; Fig. 2 of their study) using Newick notation. The resulting tree was as follows:

```
((((((Acidobacteria,Deltaproteobacteria),Proteobacteria),((Deinococcus,Chloroflexi),Cyanobacteria),((Thermotogae,Aquificae),Fusobacteria))),(((Planctomycetes,Spirochaetes),Actinobacteria),(((Chlorobi,Bacteroidetes),Fibrobacteres),Chlamydiae))),Firmicutes),(Archaea,Eukarya));
```

2.3.3. Phylogeny reconciliation

The ExbBD phylogeny was reconciled with the species tree through T-REX's horizontal gene transfer detection method (Alix et al., 2012) and Notung 2.6 (Vernot et al., 2008). Notung 2.6 accomplishes tree reconciliation by inferring gene duplications and losses. In both T-REX and Notung, the ExbBD phylogeny was rooted through the midpoint method.

In T-REX, the HGT detection mode was set to "Several HGTs by iteration," and the bipartition dissimilarity optimization criterion was selected. In Notung, after reconciliation was performed, weak edges (branches with low bootstrap support) were rearranged to produce a more parsimonious scenario of gene duplications and losses.

3. Results

3.1. Phylogeny of ExbB/MotA

The ML phylogeny of ExbB/MotA sequences is shown in Fig. 1. According to this tree, ExbB and MotA form distinct monophyletic groups, and neither are directly derived from the other. The ExbB and MotA clades are well supported, with bootstrap values of 85%.

A feature of this molecular phylogeny is that many of the internal branches in the tree are not supported by statistically significant bootstrap values (>75%). Thus, while the phylogeny is useful for exploring the phylogenetic relationship of ExbB and MotA, it most probably does not reflect the actual branching order of individual lineages.

The tree rooted through the molecular clock analysis (Fig. 2) suggests that the common ancestor of the ExbB clade existed prior to the common ancestor of MotA (ExbB node age is 1.2173, 95% highest posterior density = 0.9577, 1.4769; MotA node age is 1.0396, 95% HPD = 0.7999, 1.2714; node ages are not shown in Fig. 2 and are based on clock-like substitutions per site). Like the ML tree, this phylogeny shows that ExbB and MotA are separated into two monophyletic groups, each group having descended from a common ancestor. However, while the MotA clade is supported by statistically significant Bayesian posterior probabilities (0.99 in this phylogeny), the ExbB group lacks significant support.

3.2. Phylogeny of ExbD/MotB

Midpoint and clock analysis rooting of ExbD/MotB phylogenies reveal that these proteins – like ExbB/MotA – are divided into distinct monophyletic groups (Figs. 3 and 4).

In both ExbD/MotB phylogenies, there is strong support for the monophyly of ExbD and MotB. The monophyly of ExbD and MotB in the ML tree has statistically significant bootstrap values (100%), and the tree inferred through clock analyses supports the monophyly of ExbD and MotB with Bayesian posterior probabilities of 98% and 100%, respectively.

There is one characteristic common to all four phylogenies. The overall branching orders of the bacteria phyla are not, on the whole, consistent with the proposed phylogenies for bacteria phyla (e.g., Ciccarelli et al., 2006). This may be due to the absence of significant bootstrap support for many of the internal nodes of the phylogenies, which can be explained by the presence of sequences with very limited similarity. It is known that it in large trees composed of sequences with low degrees of similarity, strong bootstrap support is not easy to obtain (Abby and Rocha, 2012). The percent identity shared between ExbB/MotA and ExbD/MotB is below the "twilight zone" (approximately 25% sequence identity; Bhardwaj et al., 2012). At this level of sequence divergence, multiple sequence alignment (MSA) and phylogeny reconstruction becomes problematic (Bhardwaj et al., 2012), and thus a lack of clear bootstrap support for many internal nodes can be expected.

3.3. Phylogeny of ExbBD

The molecular phylogenies described in Sections 3.1 and 3.2, then, while useful for understanding the relationship of the TonB components with the flagellar motor proteins, do not shed much light on the biological history of ExbB and ExbD. As such, a

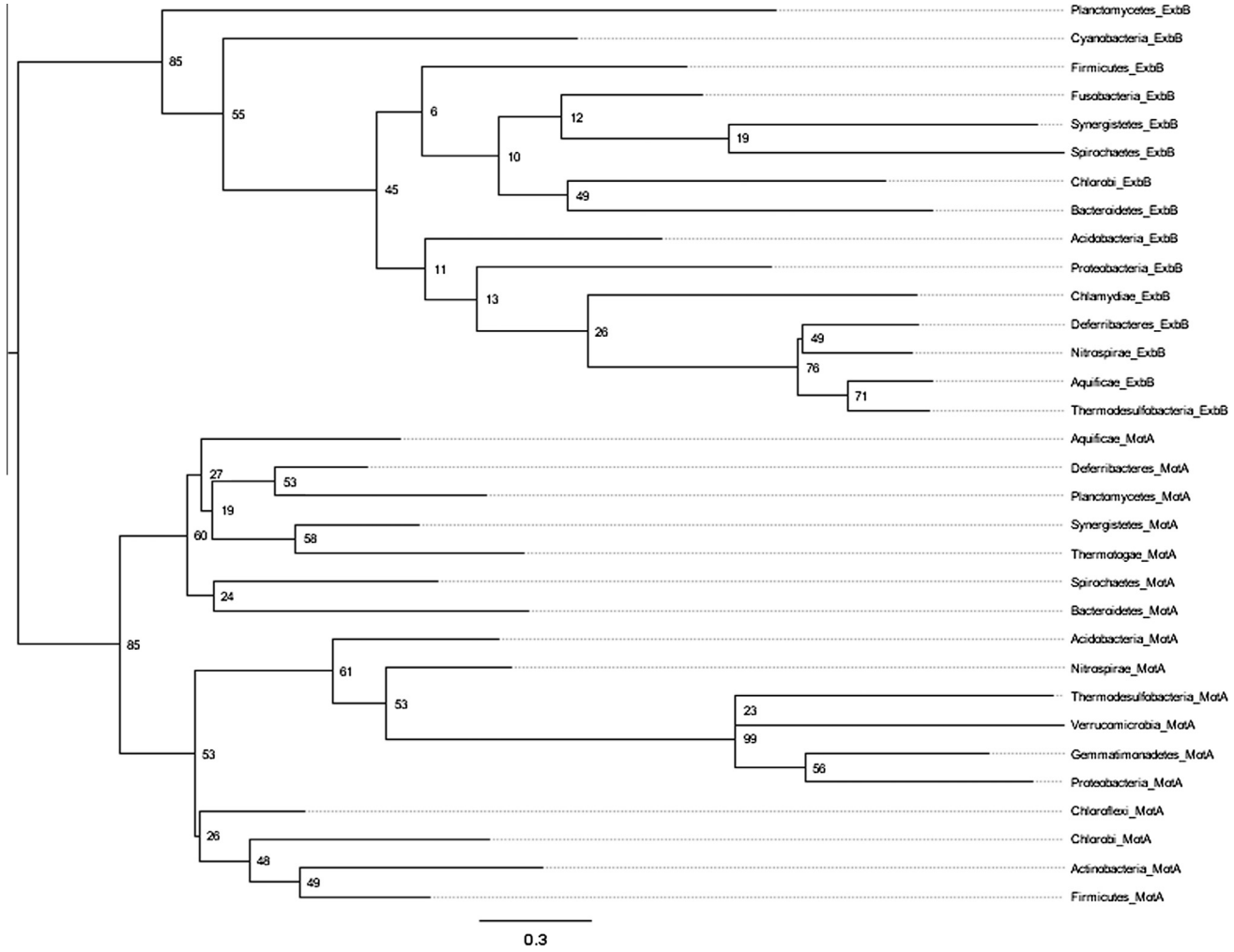

Fig. 1. Maximum-liklihood phylogeny of ExbB/MotA with bootstrap values. The phylogeny was rooted through the midpoint method.

molecular phylogeny of concatenated ExbBD sequences was estimated through maximum-likelihood and rooted with the midpoint method as detailed in Section 2.2 (see Fig. 5).

Again, this phylogeny is not consistent with previously published trees of bacteria phyla. Thus, a reconciliation approach was warranted, wherein the molecular phylogeny of ExbBD was mapped onto the species tree of Ciccarelli et al. (2006). By so doing, inferences regarding the evolutionary history of ExbBD could be made. Reconciliation was carried out through two scenarios: gene duplication/losses and horizontal gene transfers.

When the ExbBD tree was compared to the species tree in Notung (and when weak edges were rearranged), it was observed that 1 duplication and 5 losses were needed to reconcile the trees (see Fig. 6). According to this scenario, a duplication of ExbB and ExbD genes early in the history of the Bacteria lead to sequence divergence of the duplicated genes. Subsequently, several losses occurred; specifically, in the Firmicutes and Cyanobacteria there was a loss of ExbB and ExbD paralogs.

Also, losses occurred in the ancestor of the Deinococcus–Thermus/Chloroflexi clade, the Fusobacteria/Aquificae/Thermotogae clade, and the Proteobacteria/Acidobacteria group. An alternative to a model consisting of gene duplications and losses is one that invokes horizontal gene transfer (HGT) events. There is ample evidence that horizontal gene transfer has played a major role in driving the diversity of prokaryotic genomes (e.g., Treangen and Rocha, 2011; Bansal et al., 2013), and such transfers often cause phylogenetic incongruences (Galtier and Daubin, 2008). By taking HGT into consideration, it is possible to reconcile conflicting phylogenies (e.g., Boussau et al., 2008), and this permitted the reconciliation of the ExbBD phylogeny with the bacteria tree (Fig. 7).

A total of 6 transfers are necessary to reconcile the ExbBD tree to the species tree. Transfers from the Firmicutes to the Cyanobacteria, from Fusobacteria to Bacteroidetes, from Proteobacteria to Chlamydiae, and from Aquificae to the Spirochaetes summarize the HGT events involving extant taxa. Two transfers to internal nodes (see Fig. 7) also occurred under this model. The gene transfer scenario in Fig. 7 allows us to make several inferences regarding the order of appearance of ExbBD in bacteria phyla. For example, it may be deduced that the ExbBD genes in Aquificae arose before those genes in Spirochaetes since the Spirochaetes acquired ExbBD genes from Aquificae. The ExbBD genes in the Firmicutes were present before those genes were in Cyanobacteria, Planctomycetes, Spirochaetes, Chlaymidae, Chlorobi, and Bacteroidetes. Indeed, if we accept this scenario of reconciliation, it is clear that the ExbBD genes in the Firmicutes represent the most basal extant lineage. Since no ExbBD genes were apparently transferred to the Firmicutes, and since the Firmicutes are at the earliest division of the Bacteria (Ciccarelli et al., 2006), the HGT model suggests that the ExbBD proteins in Firmicutes are the deepest-branching lineage of the OTUs.

There are two competing schemes that would reconcile the ExbBD tree to the species tree: gene duplication/loss and horizontal

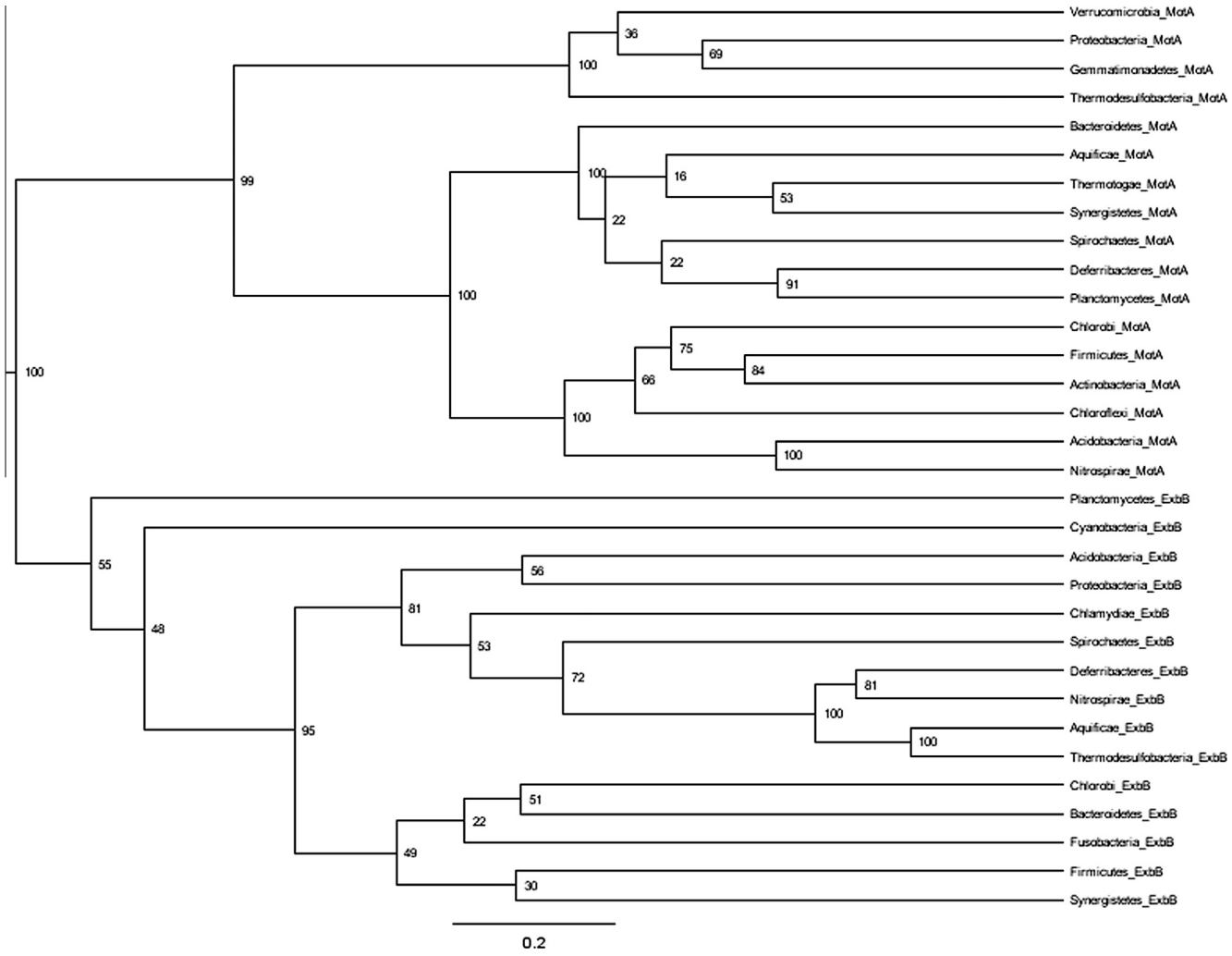

Fig. 2. A tree of ExbB/MotA generated through Bayesian inference and rooted through clock analysis. Bayesian posterior probabilities are shown on the nodes.

gene transfer. Which one is most likely to be a more accurate reflection of biological history? Both are equally parsimonious, with each requiring a total of six steps. Bioinformatic approaches can help in detecting laterally transferred genes. Among these approaches is that of comparing the GC-content of the gene with that of the genome (Zaneveld et al., 2008), as was described in Section (2.3.1). The results of this analysis yielded some interesting discoveries (see Tables 4 and 5).

Importantly, there is evidence from GC-content analysis that the Spirochaete ExbD gene was acquired from the Aquificae genome because both of these have equal GC-content values. This is entirely consistent with the model of horizontal gene transfer displayed in Fig. 7. Furthermore, the GC-content of the Cyanobacteria ExbB is quite different from its genome, suggesting an HGT event. Interestingly, with both ExbB and ExbD, the greatest differences in GC-content lie in those ExbB and ExbD genes which are proposed to have been horizontally transferred. Also, the mean difference in GC-content between genes hypothesized to have been laterally acquired and their genomes is greater than the mean difference between vertically transferred genes and their genomes.

An independent-samples *t*-test was carried out to compare the differences between the means of these two datasets: firstly, a dataset consisting of the differences in GC-content between genomes of phyla that did not acquire ExbBD laterally and their ExbBD genes, and secondly a dataset of the differences between GC-content in the genomes of phyla that are hypothesized to have acquired ExbBD laterally and their ExbBD genes. There was a significant difference ($\alpha = 0.05$) between the mean of the first dataset ($M = 1.875$, SD = 1.5) and the mean of the second dataset ($M = 3.5$, SD = 2.2038); $t(22) = 2.1385$, $p = 0.0438$. This is statistical evidence that GC-content analysis is in accord with the HGT model.

Certainly, it is possible that a combination of both HGT and gene duplication/loss of ExbBD genes has occurred. Nevertheless, a simple model of HGT is capable of reconciling the ExbBD tree with the species tree, and this model is supported by analysis of the GC-content of genes and genomes. Thus, in Section 4 it will be assumed that the HGT model is the better reconciliation framework and that the ExbBD complex in Firmicutes represents the most basal extant lineage of these proteins.

4. Discussion

4.1. Ancient origins of the ExbBD complex

It is possible to make a rough estimate of when the ExbBD complex – and more broadly, the TonB system – first originated. It is known that the photosynthetic apparatus of Cyanobacteria requires iron atoms in order to fully function (Ferreira and Straus, 1994; Mirus et al., 2009), and iron starvation may result in structural changes in cyanobacterial cells (Wang et al., 2010). Uptake

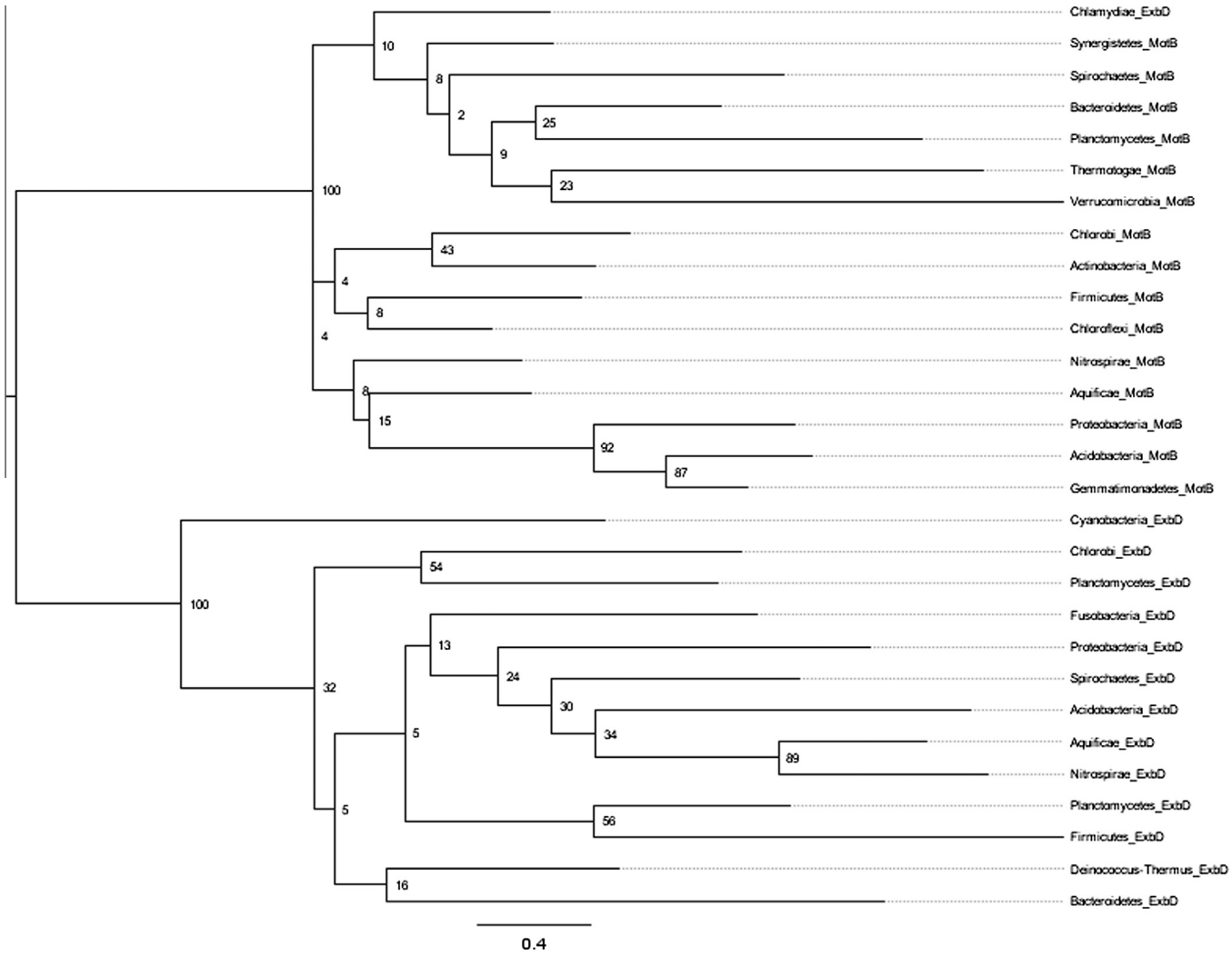

Fig. 3. Phylogeny of ExbD/MotB as estimated through the ML method, with bootstrap values next to the nodes.

of iron ions in cyanobacteria must be highly regulated to prevent molecular damage from oxidation (Mirus et al., 2009). These observations would seem to place the origin of the TonB–ExbB–ExbD complex prior to the rise of photosynthesis in the ancestor of Cyanobacteria, since it is unlikely that the photosynthetic machinery could arise and function without the presence of an effective catalyst for iron transport. Evidence indicates that photosynthesis first evolved in the cyanobacterial lineage (Mulkidjanian et al., 2006), and that photosynthesis arose well before the 'Great Oxidation Event' (Buick, 2008) which is thought to have occurred between ~2.45 Ga and ~2.32 Ga (Farquhar et al., 2011). All these considerations date the origin of the TonB–ExbB–ExbD complex as prior to ~2.32 Ga. – ~2.45 Ga. Further, there is evidence for oxygenic photosynthesis around 2.72 Ga, and even as far back as 3.2 Ga (Buick, 2008). Thus, the TonB–ExbB–ExbD complex may have originated as far back as 3 billion years ago, and quite possibly before the 'Great Oxidation Event'.

Indeed, the molecular phylogeny of ExbBD implies that this complex existed at the time of the split between Firmicutes and the other bacteria phyla. So the ExbBD complex was probably present in the Last Common Ancestor (LCA) of the Bacteria. ExbB and ExbD are also found in Archaea, hinting at an even earlier origin for ExbBD. However, the ExbBD protein sequences are only found in one phylum of Archaea, the Euryarchaeota. When the NCBI protein sequence database is used to search for ExbBD proteins in other Archaea phyla, no ExbBD proteins are found – evidence that the ExbBD proteins were transferred horizontally after Archaea had originated, and were transferred only to one phylum.

4.2. The TonB system and the flagellar motor

The ExbBD and MotAB proteins constitute unique monophyletic clades which are derived from a common root, and neither complex appears to be directly derived from the other. The present Section (4.2) will be limited to a discussion of this question: which came first, the ExbBD complex or the flagellar motor?

In general, the branch lengths of the ExbBD clade in the phylogenies are longer than those of the MotAB clade (see, e.g., Figs. 2 and 4, which reports the node ages of the ExbB/MotA and ExbD/MotB clades; node ages represent substitutions per site), which implies that they are more diverse than MotAB. This in turn suggests that ExbBD is more ancient than MotAB, assuming a near-constant molecular clock. However, the possibility that ExbBD has undergone a more rapid rate of molecular evolution relative to MotAB cannot be dismissed. The TonB system is involved in virulence in various pathogens (e.g., Beddek et al., 2004), and virulence systems may evolve more rapidly than non-virulence systems. Nevertheless, the molecular phylogenies of these proteins must be taken as evidence for the view that ExbBD arose first. Evidence has also been presented in Section 4.1 that the TonB complex must have

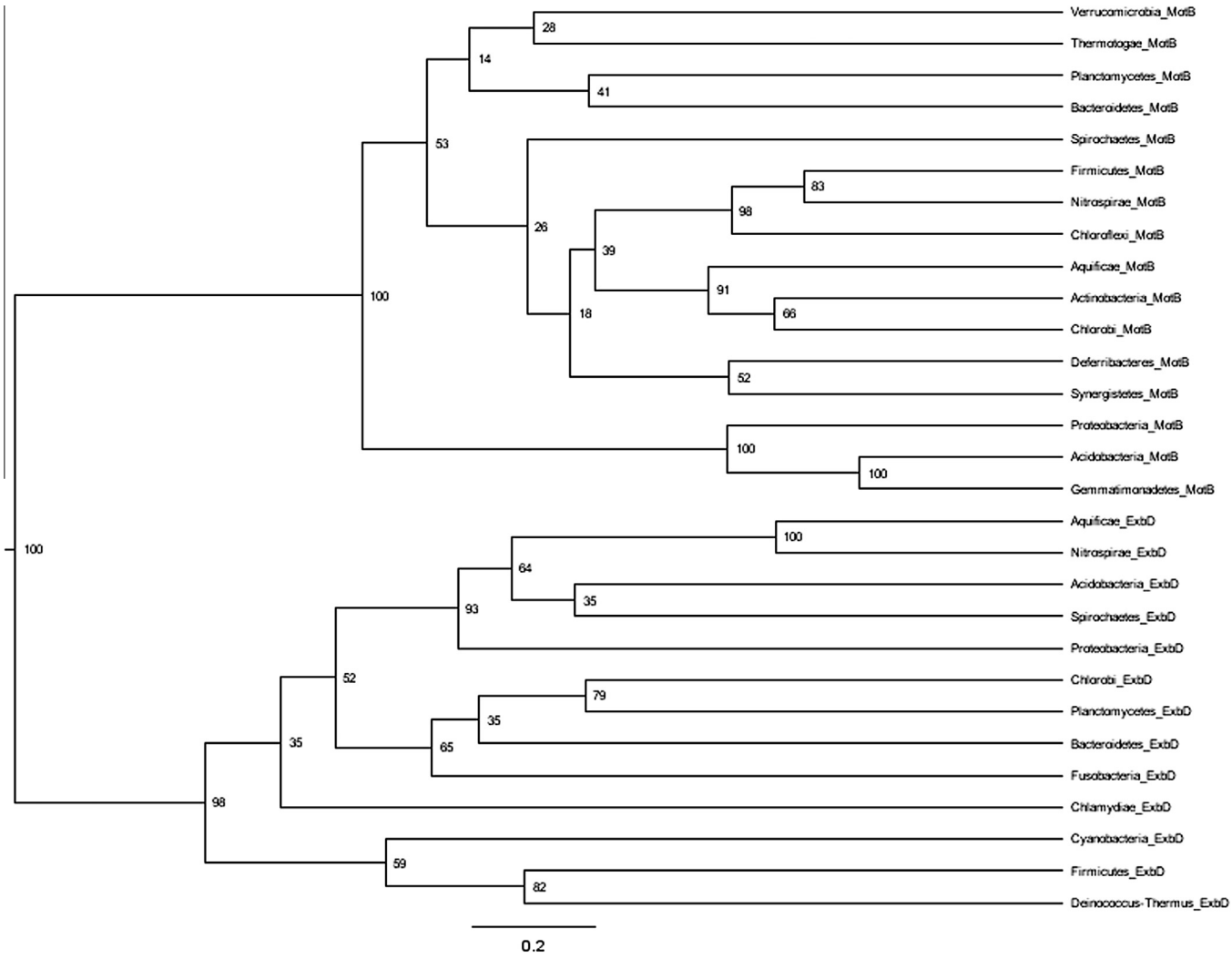

Fig. 4. A phylogeny estimation of ExbD/MotB rooted through clock analysis. Node numbers are Bayesian posterior probabilities. The node age (not shown in figure) of the ExbD clade is 1.4001 (95% HPD = 1.1533, 1.67), and the node age of MotB is 1.1432 (95% HPD = 0.9351, 1.3614).

originated before ~2.5 billion years ago, at minimum. Thus, if the flagellar motor complex is more ancient than the TonB complex, it must have originated prior to oxygenic photosynthesis.

Is there evidence that the bacterial flagellum existed before the TonB system? The flagellar ATP synthase provides evidence that a primal flagellar system existed in the last universal common ancestor (LUCA). The FliI component of the flagellar ATP synthase is homologous to the alpha and beta subunits of the F1F0-ATPase (Albertini et al., 1991), and shares closer similarity to the beta subunit than to the alpha subunit (Vogler et al., 1991). Gogarten and Kibak (1992) analyzed the phylogenetic relationships of VFA-ATPase subunits, as well as FliI, and concluded that FliI was present in the last universal common ancestor (LUCA). The flagellar basal body possesses other proteins homologous to F0F1-ATPase, namely FliH (Pallen and Matzke, 2006) and FliJ (Ibuki et al., 2011). Scenarios for the origin of the bacterial flagellum typically involve the *in toto* co-option of an ATPase (e.g., Matzke, 2003), so if FliI was in the LUCA, then so too was FliH and FliJ. This, then, is evidence that at least some of the core flagellar proteins were already possessed by the LUCA. Furthermore, it lends some plausibility to the notion that the LUCA possessed a primitive flagellar system, and this raises the possibility that the flagellar motor is more ancient than the ExbBD complex. It should be pointed out that similar evidence for the presence of the TonB system in the LUCA is lacking.

Finally, the answer to the question of which system is more ancient hinges on whether gram-positive bacteria are phylogenetically basal or not. The TonB and flagellar systems carry out different functions: the TonB system acts as a catalyst for active transport in gram-negative bacteria, while the flagellar system provides motility for both gram-negative and gram-positive bacteria. Thus if gram-positive cells originated first, it is probable that the flagellar system originated prior to the TonB system (and thus the MotAB complex is more ancient than ExbBD), given that the TonB system is not known to function in a gram-positive context (though gram-positive bacteria do harbor ExbBD and TonB genes, suggesting at least some kind of function). On the other hand, if gram-negative bacteria are phylogenetically more basal, the TonB system is likely to be more ancient than the flagellar system – particularly since the TonB function is probably much more necessary for gram-negative bacteria than motility. Thus, to a large extent the question of which system is more ancient hinges on whether gram-positive bacteria or gram-negative bacteria arose first. I offer the above considerations for the reader's benefit; future studies, perhaps utilizing molecular clocks, can definitely demonstrate which system is older.

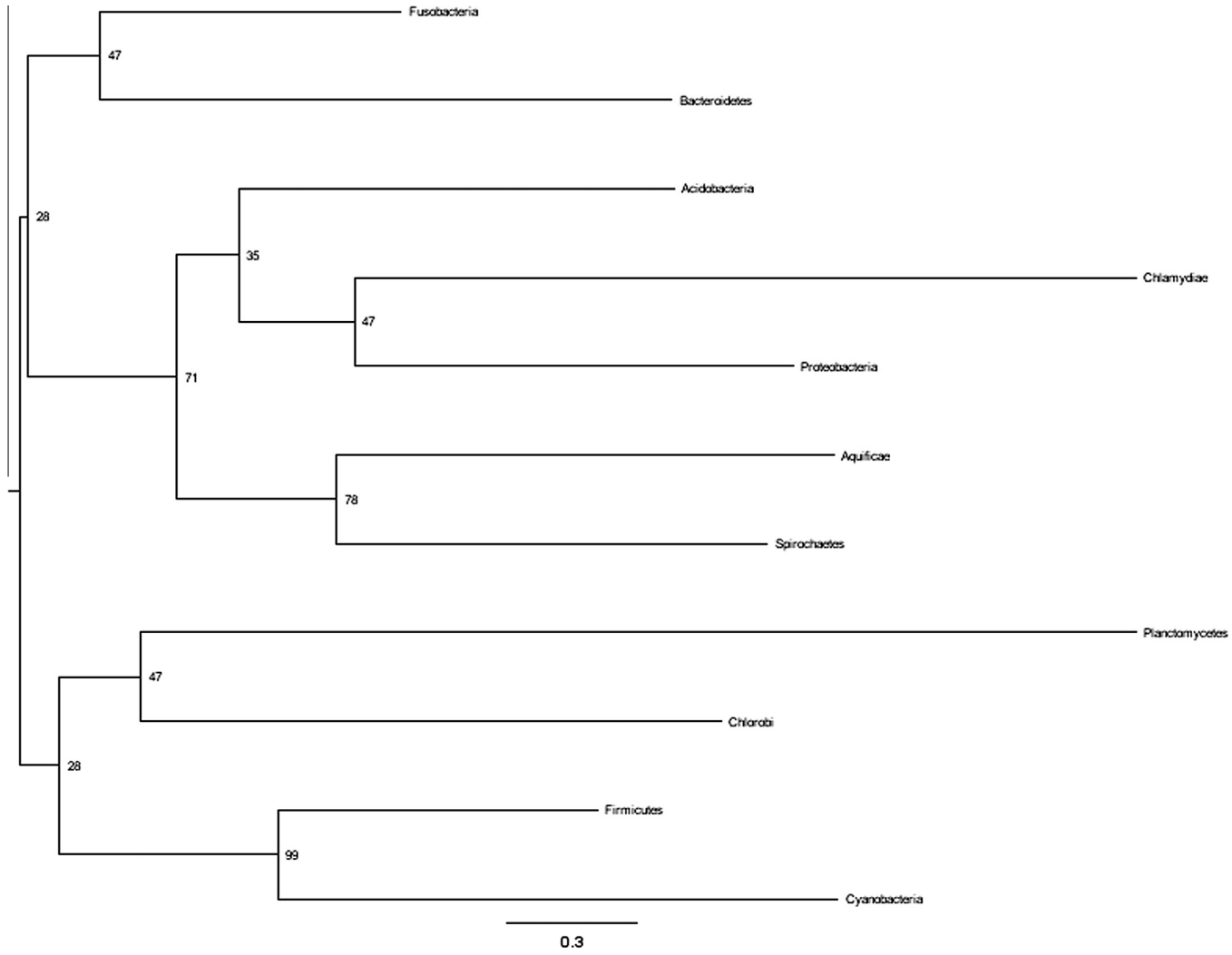

Fig. 5. A phylogeny generated from concatenated ExbB and ExbD sequences. Bootstrap values depicted next to the nodes.

4.3. The evolution of the TonB system

When it comes to the evolution of the TonB system, three scenarios present themselves: firstly, ExbB and ExbD associated with the TonB protein separately; secondly that ExbBD first arose as a proton channel and then interacted with TonB as a whole; and thirdly that ExbBD evolved directly from the flagellar motor.

4.3.1. The "TonB-early" hypothesis

This model proposes that TonB, ExbB, and ExbD were originally functioning in contexts independent of each other. A number of functions could be postulated for a proto-TonB. For example, a proto-TonB might have been involved in interactions with outer membrane pores, perhaps simply opening gated pores. A primitive TonB homolog need not be in association with ExbB–ExbD in order to function in this context, demonstrated by the observation that a TonB-like protein and an ATP-binding cassette protein function together in opening an OM secretion channel (Howard et al., 1996). Next, ExbB may have interacted with TonB, providing a stabilizing role of some kind. Again, this step is not biologically unrealistic, as it has been proposed that ExbB stabilizes TonB (Pramanik et al., 2010). Association of ExbD with the primitive TonB–ExbB complex would result in a fully functional TonB system. Finally, duplication of ExbB and ExbD would produce a proton channel that would be subsequently integrated into a primitive flagellar system, explaining the origin of MotA and MotB.

This model for TonB system evolution suffers from a certain difficulty, however. If MotAB evolved directly from the ExbBD system, then it is logical to expect that MotAB sequences would nest within ExbBD. However, the molecular phylogenies of these protein sequences show no such nesting.

4.3.2. The "proto-ExbBD proton channel" model

Another hypothesis for the origin of the TonB system is as follows: a proto-ExbB and proto-ExbD associated and functioned as a simple proton channel which would permit the regulation of hydrogen ion flow across the cell membrane, balancing pH levels in the cytoplasm. A TonB copy would then associate with this proton channel, giving rise to the TonB system. The flagellar motor would arise through co-option of this proton channel for motility purposes. This hypothesis is the most direct and parsimonious; it explains the homology shared between ExbB/MotA and ExbD/MotB. It is also in complete accord with the phylogenies described in this study, which suggest that ExbB/MotA and ExbD/MotB descended from a common ancestral system.

4.3.3. Evolution from the flagellar motor

The ExbBD complex may have evolved from the flagellar motor proteins in a stepwise fashion which would involve duplication of MotA and MotB (forming a proton channel), followed by TonB binding to the ExbBD proton channel. Such a hypothesis implies that a primitive flagellum must have existed earlier than the TonB

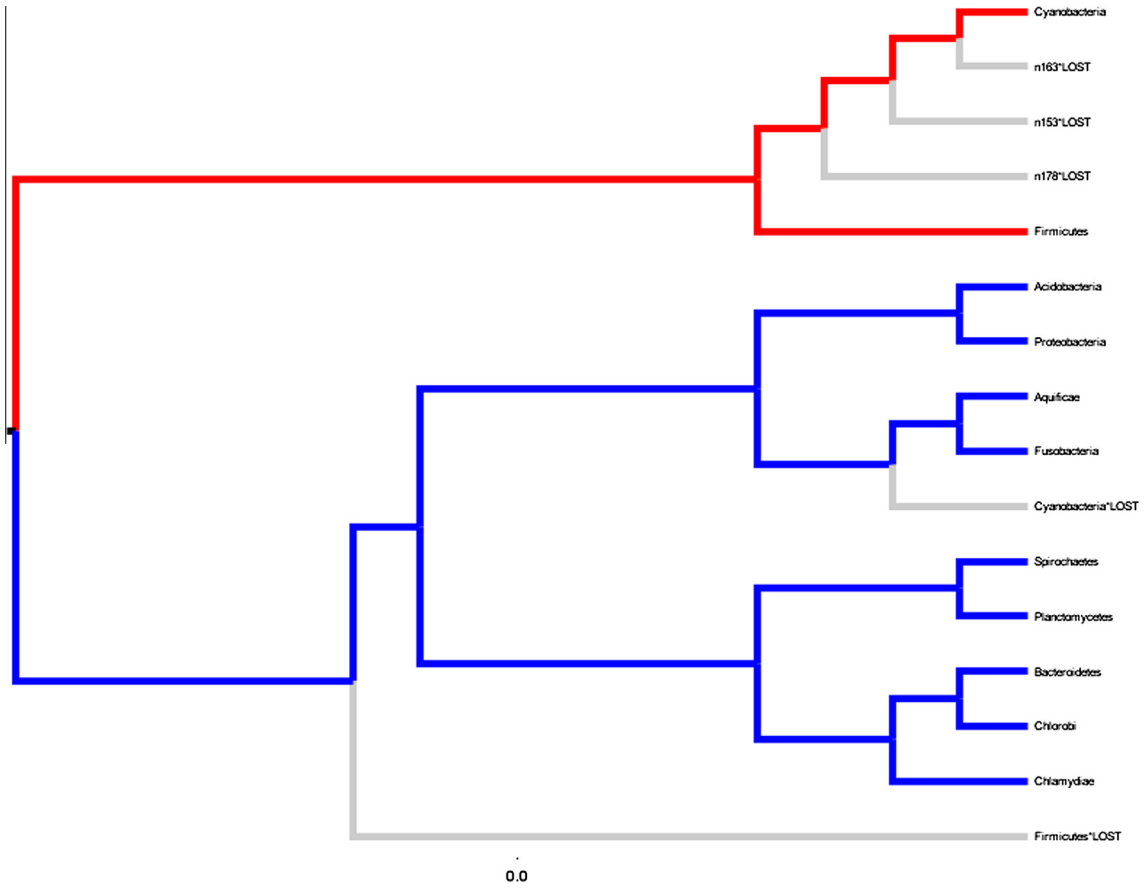

Fig. 6. The ExbBD phylogeny reconciled to a species tree of bacteria phyla. Branches in blue and in red represent different gene lineages. Grayed-out branches illustrate losses. The duplication/loss (D/L) score is 6.5. This reconciliation approach begins with a duplication of ExbBD genes in an ancestor of the Firmicutes. One of the ExbBD duplicates was later lost in the Firmicutes, and losses of ExbBD duplicates also occurred in several other taxa, such as the Cyanobacteria. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

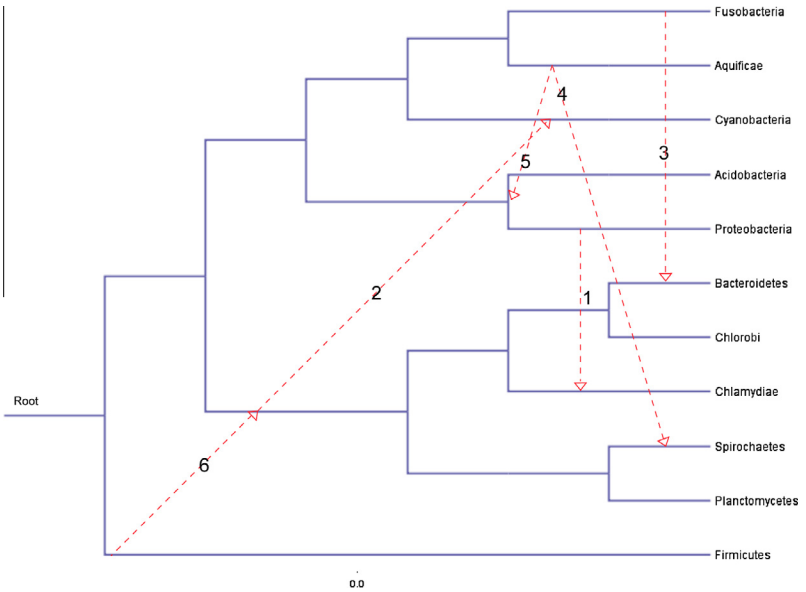

Fig. 7. A phylogeny depicting the various horizontal gene transfers required to reconcile the ExbBD tree to the species tree. The cladogram shown here is the species tree; the arrows represent HGTs and the direction of transfer.

system. The antiquity of the flagellar ATPase attests to the possibility that the flagellum predates the TonB system, so from this perspective it is not unreasonable to propose that the ExbBD proteins are descendants of MotAB.

4.3.4. Analysis of the models

Of all three hypotheses for the origin of the TonB system, the first is not likely to be correct for the reason explained in Section 4.3.1. The thesis that the common ancestral system was a simple

Table 4
GC-content of the ExbB gene from a given phylum (right column) and the GC-content of the genome in which the gene was located (left column). Asterisks denote phyla which are hypothesized to have laterally acquired an ExbB gene, based on Fig. 7.

Bacteria phyla	Genome GC-content (%)	ExbB GC-content (%)
Acidobacteria	64	64
Aquificae	43	43
Bacteroidetes*	43	40
Chlamydiae*	41	40
Chlorobi	53	54
Cyanobacteria*	48	53
Firmicutes	36	40
Fusobacteria	34	37
Nitrospirae	59	55
Planctomycetes	62	66
Proteobacteria	54	58
Spirochaetes*	35	38

Table 5
GC-content of the ExbD gene from a given phylum (right column) and the GC-content of the genome in which the gene was located (left column). Asterisks denote phyla which are hypothesized to have laterally acquired an ExbD gene, based on Fig. 7.

Bacteria phyla	Genome GC-content (%)	ExbD GC-content (%)
Acidobacteria	58	57
Aquificae	43	42
Bacteroidetes*	48	50
Chlamydiae*	41	43
Chlorobi	53	52
Cyanobacteria*	47	51
Firmicutes	56	58
Fusobacteria	27	26
Nitrospirae	59	59
Planctomycetes	55	57
Proteobacteria	54	52
Spirochaetes*	35	43

proton channel seems to this author to be more satisfactory than the idea that ExbBD evolved directly from the flagellar motor. Not only does it predict the observed tree topologies depicted in Figs. 1–4, but it is also more parsimonious: fewer steps are needed to explain origin of the TonB system and flagellar motor. Evolution of ExbBD from the flagellar motor would demand the following steps: the evolution of the flagellar motor from some remote homologs, duplication of the flagellar motor to form an intermediate proton channel, and association of TonB with this proton channel, giving rise to the TonB system. On the other hand, if the ExbBD and MotAB complexes evolved from a proton channel that predated either one, only two steps are required: interaction of TonB with the proton channel and co-option of the proton channel for flagellar function, resulting in MotAB. Nonetheless, caution should be used in reaching this conclusion as it has been argued that unweighted parsimony – wherein gains and losses do not have equal weight – is an unreliable method to the reconstruction of phylogeny (Cavalier-Smith, 2002). Only further research will be able to definitely resolve the question of how the TonB system evolved, but at present the "proto-ExbBD proton channel" model is the most strongly supported by the evidence.

4.4. A timeline of the biological history of the TonB system

The ExbBD sequences in the Firmicutes represent the deepest-branching extant lineage of this protein complex. Insight on the biological history of the TonB system might be gained by looking within the Firmicutes. The Firmicutes are primarily gram-positive (Vollmer, 2011), with the exception of a few taxa like the Negativicutes, which stain gram-negative (Marchandin et al., 2010). The genomes of some gram-positive Firmicute taxa encode TonB, ExbB, and ExbD proteins. ExbD is only known to be present in *Halothermothrix orenii* (Mavromatis et al., 2009) and gram-negative Firmicutes; ExbB has a slightly wider distribution in this phylum, and proteins partially similar to TonB are scattered throughout the Firmicutes. The TonB–ExbB–ExbD operon in *Halothermothrix orenii,* then, appears to be the only operon of this kind within the gram-positive Firmicutes. That is, in the gram-positive Firmicutes, the *Halothermothrix orenii* are the sole bacteria that possess all three proteins required for the TonB system. However, the *Halothermothrix orenii* have both gram-positive and gram-negative features, and are located within a gram-negative order, so it is not entirely accurate to describe them as gram-positive. Intriguingly, no TonB-dependent receptors have been detected in *Halothermothrix orenii* (Mavromatis et al., 2009), suggesting that the TonB–ExbB–ExbD complex in this taxon has a different function than that of other TonB systems. TonB-dependent receptors are present in some gram-negative Firmicutes, though. For example, the TonB-dependent receptor FepA (Ferguson and Deisenhofer, 2002) has unambiguous homologs in the Negativicutes and some Halanaerobiales strains (both gram-negative). These observations suggest a chronology for the origin of the TonB system: TonB, ExbB, and ExbD first merged together in a species reminiscent of the *Halothermothrix orenii*. Possibly, the TonB system of *Halothermothrix orenii* is a descendent of this primitive TonB system. I hypothesize that like the *Halothermothrix orenii,* this species had both gram-positive and gram-negative properties, and lacked TonB-dependent receptors. Consequently, this early TonB system probably possessed a primitive membrane function not completely related to the transport of iron ions. With the rise of full-fledged gram-negative Firmicutes came the origin of outer membrane receptors like FepA. These receptors shifted the function of the primal TonB system to the more specific task of iron transport. So the TonB system probably pre-dates outer membrane receptors. Next, lateral and vertical transfer of ExbBD genes occurred; the Spirochaete ExbBD complex, for example, arrived from the Aquificae genome. So, too, was the ExbBD complex of the Cyanobacteria inherited laterally – in this instance, from the Firmicutes. Adaptation to specific niches is also part of the evolutionary history of the TonB system: the TonB2 and TonB3 systems of *Vibrio vulnificus* possess a fourth protein known as TtpC (Kustusch et al., 2012). Thus, the biological history of the TonB system is one of co-option, horizontal gene transfer, and adaptation and diversification.

4.5. Research on the TonB system

This study raises several questions and opens the door for further research. The focus here has been on the ExbBD components of the TonB system; a molecular phylogenetic analysis of TonB protein sequences across prokaryotic phyla has yet to be conducted. Such work would answer various questions: have TonB protein sequences been laterally transferred in conjunction with ExbBD? Does the overall phylogeny of TonB concord with that of ExbBD? Molecular clock analyses, too, would be useful in determining the divergence times of TonB, ExbB, and ExbD. These analyses could also be applied to both the TonB system and the flagellar motor to discover when these two systems diverged.

In addition, the results in this study may help with resolving prokaryotic phylogeny. If the HGT transfer scenario presented here is indeed correct, then this is evidence for the validity of the Ciccarelli et al. (2006) branching order. Use of a different species tree would lead to a different scheme of HGT, and by determining which scheme is correct it should be possible to test the accuracy of each published species tree. Here, too, molecular clocks can be harnessed for this task. For example, the HGT scenario of ExbBD in this study predicts that the Cyanobacteria ExbBD protein complex is less ancient than the ExbBD complex in the Firmicutes (since the ExbBD complex of Cyanobacteria was transferred from

the Firmicutes). By examining the ExbBD protein sequences within the Cyanobacteria and Firmicutes, this prediction can be tested.

5. Conclusions

Phylogenetic analyses of the ExbBD components of the TonB system and the MotAB proteins of the flagellar motor suggest that ExbB/MotA and ExbD/MotB are both divided into different monophyletic groups having unique evolutionary histories. MotAB did not evolve from ExbBD, and ExbBD did not descend from MotAB. The branch lengths of ExbBD are typically longer than those of MotAB, indicating that there has been greater sequence divergence among ExbB and ExbD sequences. Furthermore, it is argued here – based on the phylogeny reconciliation of ExbBD and GC-content analysis – that these proteins have been laterally transferred throughout bacteria phyla several times. It is suggested that the TonB system first arose in an organism with gram-negative characteristics, that this was followed by the origin of outer membrane receptors, and that the TonB system subsequently diversified. The evolution of the TonB system can be summarized in this manner:

(a) A proto-ExbB bonded to a proto-ExbD, forming a basic proton channel. The original functions of either of these proteins are somewhat difficult to envision, particularly given the absence of distant homologs other than MotA and MotB.
(b) A TonB-like protein, probably acting as a channel opener, then interacted with the proto-ExbBD complex, yielding a primitive TonB system.
(c) The modern TonB system evolved through optimization of this early machine over evolutionary time.

This study presents several avenues for future research. Dating the origin of either ExbB/ExbD or MotA/MotB can be accomplished using molecular clocks, and this will help in rigorously dating the origin of these protein complexes. This can then be correlated with other events in biological history to determine the impact that the origin of the TonB and flagellar systems had in shaping the course of prokaryotic evolution. Also, the phylogenetic relationships of the TonB and Tol-Pal complexes have yet to be fully understood, and thus research may be focused in this area.

In conclusion, inferring the evolutionary and functional relationships of molecular machines will aid in deciphering their origin, and, to that end, the phylogenetic reconstructions of protein components advance our understanding of the origin of molecular machines.

Appendix A

The accession numbers of the protein sequences used in this study are listed in Tables 6.1–6.4.

Table 6.1
ExbB accession numbers.

Acidobacteria	ZP_09578070.1
Aquificae	O67637.1
Bacteroidetes	CAZ96486.1
Chlamydiae	NP_877085.1
Chlorobi	ABP37390.1
Cyanobacteria	BAL31299.1
Deferribacteres	YP_003503651.1
Firmicutes	ZP_08250100.1
Fusobacteria	YP_003967398.1
Nitrospirae	YP_003796720.1
Planctomycetes	ZP_09572524.1
Proteobacteria	YP_003521632.1
Spirochaetes	YP_003605.1
Synergistetes	ZP_07738809.1
Thermodesulfobacteria	YP_004627877.1

Table 6.2
ExbD accession numbers.

Acidobacteria	ABF39156.1
Aquificae	O67694.1
Bacteroidetes	YP_004510476.1
Chlamydiae	AAP98741.1
Chlorobi	ABP37388.1
Cyanobacteria	YP_001519191.1
Deinococcus–Thermus	YP_004299.1
Firmicutes	ZP_08501595.1
Fusobacteria	ZP_04970196.1
Nitrospirae	YP_003796345.1
Planctomycetes	NP_869568.1
Proteobacteria	YP_003521631.1
Spirochaetes	AAS72243.1

Table 6.3
MotA accession numbers.

Acidobacteria	YP_002753258.1
Actinobacteria	ZP_10950995.1
Aquificae	O67122.1
Bacteroidetes	YP_446692.1
Chlorobi	YP_005847478.1
Chloroflexi	YP_002523314.1
Deferribacteres	YP_003496930.1
Firmicutes	ABS73708.1
Gemmatimonadetes	YP_002759808.1
Nitrospirae	YP_003797994.1
Planctomycetes	YP_005445817.1
Proteobacteria	YP_003520527.1
Spirochaetes	AEW68621.1
Synergistetes	CBL28134.1
Thermodesulfobacteria	YP_004628544.1
Thermotogae	YP_005470316.1
Verrucomicrobia	ZP_10268475.1

Table 6.4
MotB accession numbers.

Acidobacteria	YP_002755734.1
Actinobacteria	ZP_10950996.1
Aquificae	ZP_04585048.1
Bacteroidetes	YP_003572686.1
Chlorobi	YP_005847477.1
Chloroflexi	YP_002523313.1
Deferribacteres	YP_003496929.1
Firmicutes	ABS73707.1
Gemmatimonadetes	YP_002759807.1
Nitrospirae	YP_003797995.1
Planctomycetes	YP_005445818.1
Proteobacteria	YP_003520526.1
Spirochaetes	YP_006202973.1
Synergistetes	ZP_06440416.1
Thermotogae	YP_002333858.1
Verrucomicrobia	YP_001940505.1

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.
Abby, S.S., Rocha, E.P.C., 2012. The non-flagellar Type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. PLoS Genet. 8 (9), e1002983. http://dx.doi.org/10.1371/journal.pgen.1002983.
Albertini, A.M., Caramori, T., Crabb, W.D., Scoffone, F., Galizzi, A., 1991. The flaA locus of Bacillus subtilis is part of a large operon coding for flagellar structures, motility functions, and an ATPase-like polypeptide. J. Bacteriol. 173 (11), 3573–3579.
Alix, B., Boubacar, D.A., Vladimir, M., 2012. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. Nucleic Acids Res. 40(Web Server issue), W573-9. doi: 10.1093/nar/gks485.
Bansal, M.S., Banay, G., Harlow, T.J., Gogarten, J.P., Shamir, R., 2013. Systematic inference of highways of horizontal gene transfer in prokaryotes. Bioinformatics. 29 (5), 571–579.

Baxevanis, A.D., Ouellette, B.F., 2001. Bioinformatics, second ed. Wiley-Interscience, New York.
Beddek, A.J., Sheehan, B.J., Bossé, J.T., Rycroft, A.N., Kroll, J.S., Langford, P.R., 2004. Two TonB systems in Actinobacillus pleuropneumoniae: their roles in iron acquisition and virulence. Infect Immun. 72 (2), 701–708.
Bhardwaj, G., Ko, K.D., Hong, Y., Zhang, Z., Ho, N.L., Chintapalli, S.V., Kline, L.A., et al., 2012. PHYRN: a robust method for phylogenetic analysis of highly divergent sequences. PLoS ONE 7 (4), e34261.
Blair, C., Murphy, R.W., 2011. Recent trends in molecular phylogenetic analysis: where to next? J. Hered. 102 (1), 130–138.
Boussau, B., Guéguen, L., Gouy, M., 2008. Accounting for horizontal gene transfers explains conflicting hypotheses regarding the position of aquificales in the phylogeny of Bacteria. BMC Evol. Biol. 8, 272.
Boykin, L.M., Kubatko, L.S., Lowrey, T.K., 2010. Comparison of methods for rooting phylogenetic trees: a case study using Orcuttieae (Poaceae: Chloridoideae). Mol. Phylogenet. Evol. 54, 687–700.
Braun, V., 1995. Energy-coupled transport and signal transduction through the Gram-negative outer membrane via TonB–ExbB–ExbD-dependent receptor proteins. FEMS Microbiol. Rev. 16 (4), 295–307.
Braun, V., Gaisser, S., Herrmann, C., Kampfenkel, K., Killmann, H., Traub, I., 1996. Energy-coupled transport across the outer membrane of *Escherichia coli*: ExbB binds ExbD and TonB in vitro, and leucine 132 in the periplasmic region and aspartate 25 in the transmembrane region are important for ExbD activity. J. Bacteriol. 178 (10), 2836–2845.
Buick, R., 2008. When did oxygenic photosynthesis evolve? Philos. Trans. R Soc. Lond. B Biol. Sci. 363 (1504), 2731–2743.
Cascales, E., Lloubes, R., Sturgis, J.N., 2001. The TolQ-TolR proteins energize TolA and share homologies with the flagellar motor proteins MotA–MotB. Mol. Microbiol. 42 (3), 795–807.
Cavalier-Smith, T., 2002. The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. Int. J. Syst. Evol. Microbiol. 52 (Pt 1), 7–76.
Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., Bork, P., 2006. Toward automatic reconstruction of a highly resolved tree of life. Science 311, 1283–1287.
Clote, P., Backofen, R., 2000. Computational Molecular Biology. An Introduction. Wiley, New York.
Durbin, R., Eddy, S., Krogh, A., Mitchison, G., 1998. Biological Sequence Analysis. Cambridge University Press, Cambridge.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32 (5), 1792–1797.
Farquhar, J., Zerkle, A.L., Bekker, A., 2011. Geological constraints on the origin of oxygenic photosynthesis. Photosynth. Res. 107 (1), 11–36.
Felsenstein, J., 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Methods Enzymol. 266, 418–427.
Ferguson, A.D., Deisenhofer, J., 2002. TonB-dependent receptors-structural perspectives. Biochim. Biophys. Acta 1565 (2), 318–332.
Ferreira, F., Straus, N.A., 1994. Iron deprivation in cyanobacteria. J. Appl. Phycol. 6, 199–210. http://dx.doi.org/10.1007/BF02186073.
Galtier, N., Daubin, V., 2008. Dealing with incongruence in phylogenomic analyses. Philos. Trans. R Soc. Lond. B Biol. Sci. 363 (1512), 4023–4029.
Ghosh, J., Postle, K., 2005. Disulphide trapping of an in vivo energy-dependent conformation of *Escherichia coli* TonB protein. Mol. Microbiol. 55 (1), 276–288.
Gogarten, J.P., Kibak, H., 1992. The Bionergetics of the Last Common Ancestor and the Origin of the Eukaryotic Endomembrane System. In: Hartman, H., Matsuno K. (Eds.), The Origins and Evolution of the Cell. World Scientific Publ. Co., pp. 131–154.
Graur, D., Li, W.H., 2000. Fundamentals of Molecular Evolution, second ed. Sinauer Associates, Sunderland, MA.
Gresock, M.G., Savenkova, M.I., Larsen, R.A., Ollis, A.A., Postle, K., 2011. Death of the TonB Shuttle Hypothesis. Front. Microbiol. 2, 206.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59 (3), 307–321.
Gupta, R.S., 2001. The branching order and phylogenetic placement of species from completed bacterial genomes, based on conserved indels found in various proteins. Int. Microbiol. 4 (4), 187–202.
Hannavy, K., Barr, G.C., Dorman, C.J., Adamson, J., Mazengera, L.R., Gallagher, M.P., Evans, J.S., et al., 1990. TonB protein of Salmonella typhimurium. A model for signal transduction between membranes. J. Mol. Biol. 216 (4), 897–910.
Hess, P.N., Russo, C.A., 2007. An empirical test of the midpoint rooting method. Biol. J. Linn. Soc. 92, 669–674.
Howard, S.P., Meiklejohn, H.G., Shivak, D., Jahagirdar, R., 1996. A TonB-like protein and a novel membrane protein containing an ATP-binding cassette function together in exotoxin secretion. Mol. Microbiol. 22 (4), 595–604.
Ibuki, T., Imada, K., Minamino, T., Kato, T., Miyata, T., Namba, K., 2011. Common architecture of the flagellar type III protein export apparatus and F- and V-type ATPases. Nat. Struct. Mol. Biol. 18 (3), 277–282.
Jana, B., Manning, M., Postle, K., 2011. Mutations in the ExbB cytoplasmic carboxy terminus prevent energy-dependent interaction between the TonB and ExbD periplasmic domains. J. Bacteriol. 193 (20), 5649–5657.
Jin, G., Nakhleh, L., Snir, S., Tuller, T., 2007. Inferring phylogenetic networks by the maximum parsimony criterion: a case study. Mol. Biol. Evol. 24 (1), 324–337.
Kampfenkel, K., Braun, V., 1992. Membrane topology of the *Escherichia coli* ExbD protein. J. Bacteriol. 174 (16), 5485–5487.
Kampfenkel, K., Braun, V., 1993. Topology of the ExbB protein in the cytoplasmic membrane of *Escherichia coli*. J. Biol. Chem. 268 (8), 6050–6057.
Kojima, S., Blair, D.F., 2001. Conformational change in the stator of the bacterial flagellar motor. Biochemistry 40 (43), 13041–13050.
Krewulak, K.D., Vogel, H.J., 2011. TonB or not TonB: is that the question? Biochem. Cell Biol. 89 (2), 87–97.
Kustusch, R.J., Kuehl, C.J., Crosa, J.H., 2012. The ttpC gene is contained in two of three TonB systems in the human pathogen Vibrio vulnificus, but only one is active in iron transport and virulence. J. Bacteriol. 194 (12), 3250–3259.
Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25 (7), 1307–1320.
Lim, B.L., 2010. TonB-dependent receptors in nitrogen-fixing nodulating bacteria. Microbes. Environ. 25 (2), 67–74.
Maddison, D., Maddison, W., 2000. MacClade 4: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, MA.
Marchandin, H., Teyssier, C., Campos, J., Jean-Pierre, H., Roger, F., Gay, B., et al., 2010. Negativicoccus succinicivorans gen. nov., sp. nov., isolated from human clinical samples, emended description of the family Veillonellaceae and description of Negativicutes classis nov., Selenomonadales ord. nov. and Acidaminococcaceae fam. nov. in the bacterial phylum Firmicutes. Int. J. Syst. Evol. Microbiol. 60 (Pt 6), 1271–1279.
Matzke, N.J., 2003. Evolution in (Brownian) Space: A Model for the Origin of the Bacterial Flagellum <http://www.talkdesign.org/faqs/flagellum.html>.
Mavromatis, K., Ivanova, N., Anderson, I., Lykidis, A., Hooper, S.D., Sun, H., et al., 2009. Genome analysis of the anaerobic thermohalophilic bacterium Halothermothrix orenii. PLoS ONE 4 (1), e4192. http://dx.doi.org/10.1371/journal.pone.0004192.
Mirus, O., Strauss, S., Nicolaisen, K., von Haeseler, A., Schleiff, E., 2009. TonB-dependent transporters and their occurrence in cyanobacteria. BMC Biol. 7, 68.
Moreira, D., Philippe, H., 2000. Molecular phylogeny: pitfalls and progress. Int. Microbiol. 3 (1), 9–16.
Mulkidjanian, A.Y., Koonin, E.V., Makarova, K.S., Mekhedov, S.L., Sorokin, A., Wolf, Y.I., 2006. The cyanobacterial genome core and the origin of photosynthesis. Proc. Natl. Acad. Sci. USA 103 (35), 13126–13131.
Nei, M., 1996. Phylogenetic analysis in molecular evolutionary genetics. Annu. Rev. Genet. 30, 371–403.
Noinaj, N., Guillier, M., Barnard, T.J., Buchanan, S.K., 2010. TonB-dependent transporters: regulation, structure, and function. Annu. Rev. Microbiol. 64, 43–60.
Ollis, A.A., Postle, K., 2011. The same periplasmic ExbD residues mediate in vivo interactions between ExbD homodimers and ExbD–TonB heterodimers. J. Bacteriol. 193 (24), 6852–6863.
Ollis, A.A., Manning, M., Held, K.G., Postle, K., 2009. Cytoplasmic membrane protonmotive force energizes periplasmic interactions between ExbD and TonB. Mol. Microbiol. 73 (3), 466–481.
Pallen, M.J., Matzke, N.J., 2006. From *The Origin of Species* to the origin of bacterial flagella. Nat. Rev. Microbiol. 4 (10), 784–790.
Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T., Manuel, M., Wörheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9 (3), e1000602.
Postle, K., Kadner, R.J., 2003. Touch and go: tying TonB to transport. Mol. Microbiol. 49 (4), 869–882.
Pramanik, A., Zhang, F., Schwarz, H., Schreiber, F., Braun, V., 2010. ExbB protein in the cytoplasmic membrane of *Escherichia coli* forms a stable oligomer. Biochemistry 49 (40), 8721–8728.
Rappé, M.S., Giovannoni, S.J., 2003. The uncultured microbial majority. Annu. Rev. Microbiol. 57, 369–394.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61 (3), 539–542.
Roof, S.K., Allard, J.D., Bertrand, K.P., Postle, K., 1991. Analysis of *Escherichia coli* TonB membrane topology by use of PhoA fusions. J. Bacteriol. 173 (17), 5554–5557.
Schauer, K., Rodionov, D.A., de Reuse, H., 2008. New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? Trends Biochem. Sci. 33 (7), 330–338.
Swayne, C., Postle, K., 2011. Taking the *Escherichia coli* TonB transmembrane domain "offline"? Nonprotonatable Asn substitutes fully for TonB His20. J. Bacteriol. 193 (15), 3693–3701.
Thornton, J.W., DeSalle, R., 2000. Gene family evolution and homology: genomics meets phylogenetics. Annu. Rev. Genomics Hum. Genet. 1, 41–73.
Treangen, T.J., Rocha, E.P., 2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. PLoS Genet. 7 (1), e1001284. http://dx.doi.org/10.1371/journal.pgen.1001284.
Vernot, B., Stolzer, M., Goldman, A., Durand, D., 2008. Reconciliation with non-binary species trees. J. Comput. Biol. 15 (8), 981–1006.
Vogler, A.P., Homma, M., Irikura, V.M., Macnab, R.M., 1991. Salmonella typhimurium mutants defective in flagellar filament regrowth and sequence similarity of FliI to F0F1, vacuolar, and archaebacterial ATPase subunits. J. Bacteriol. 173 (11), 3564–3572.
Vollmer, W., 2011. Bacterial outer membrane evolution via sporulation? Nat. Chem. Biol. 8 (1), 14–18.
Wang, Q., Hall, C.L., Al-Adami, M.Z., He, Q., 2010. IsiA is required for the formation of photosystem I supercomplexes and for efficient state transition in

synechocystis PCC 6803. PLoS ONE 5 (5), e10432. http://dx.doi.org/10.1371/journal.pone.0010432.
Yang, Z., 1996. Phylogenetic analysis using parsimony and likelihood methods. J. Mol. Evol. 42 (2), 294–307.
Zaneveld, J.R., Nemergut, D.R., Knight, R., 2008. Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. Microbiology 154 (Pt 1), 1–15. http://dx.doi.org/10.1099/mic.0.2007/011833-0.
Zhai, Y.F., Heijne, W., Saier Jr., M.H., 2003. Molecular modeling of the bacterial outer membrane receptor energizer, ExbBD/TonB, based on homology with the flagellar motor. MotAB. Biochim. Biophys. Acta 1614 (2), 201–210.

Web References

NCBI protein sequence database: <http://www.ncbi.nlm.nih.gov/protein>. (accessed 2.10.12).
PhyML 3.0 web interface: <http://www.hiv.lanl.gov/content/sequence/PHYML/interface.htm>. (accessed 4.10.12).
FigTree: <http://tree.bio.ed.ac.uk/software/figtree/>. (accessed 15.10.12).
EMBOSS 6.3.1 geecee: <http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::geecee>. (accessed 12.04.13).
NCBI genome database: <http://www.ncbi.nlm.nih.gov/genome>. (accessed 12.04.13).