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