Comparative Phylogeny of the Genus *Bordetella* Using Sequence Analysis of 16S rRNA and *ompA* Genes

**Ali Badamchi¹, Moslem Papizadeh²*¹**

¹ Children’s Medical Center Hospital, Tehran University of Medical Sciences, Tehran, Iran.

² Department of Microbiology, Pasteur Institute of Iran (IPI), Tehran, Iran.

**ARTICLE INFO**

**Article type:** Original Article

**Article history:**
Received: 19 Jan 2017
Revised: Jun Mar 2017
Accepted: 11 Sep 2017
Published: 15 Oct 2017

**Keywords:** Alcaligenaceae, Biogeography, Bordetella species, Ecological distribution, Phylogenetic species concept.

**ABSTRACT**

**Background:** The genus *Bordetella* harbors 16 species; three of them are well-known for their high medical importance. The phylogenetic diversity of the genus is currently not very well investigated.

**Methods:** In this study, 16S rRNA gene sequence of 16 type strains of the *Bordetella* species were analyzed. Also, phylogenies conducted on the same gene of 247 isolates of *Bordetella* species, comprising a wide physiological as well as ecological diversity and encompassing ex-type representatives of the 16 *Bordetella* species, were analyzed.

**Results:** It was found that the phylogenetic diversity of the genus may be very different from that is currently assumed. Interestingly, the 16S rRNA gene signals could not resolve some species with promising bootstrap and posterior probability values as our phylogenies, using maximum likelihood and Bayesian inference methods, showed.

**Conclusion:** Our results indicate a probable need for additional phylogenetic signals which can be provided by coding genes. Therefore, sequence data of *ompA* gene of *Bordetella* species, a critically significant genomic region in pathogenesis, was here analyzed, phylogenetically. This gene confirmed the tree topology and the phylogenetic species boundaries already revealed by the 16S rRNA gene, but showed a better discriminatory power which resolved *Bordetella* species with higher statistically significant values.

Introduction

There is a considerable number of opportunistic bacterial pathogens in various environmental samples including soils and feces (1, 2). Various soils are known as the origin of various non-pathogenic and either pathogenic microbial species. Thus, various species of opportunistic bacterial pathogens, likely *Enterobacteriaceae* (3), *Microbacterium* (4), *Pseudomonas* (5), *Stenotrophomonas* (6), and many other genera of the kingdom *Eubacteria* can be detected in various soils, abundantly.

The genus *Bordetella* includes 16 well-known species among which three species: *B. pertussis*, *B. parapertussis*, and *B. bronchiseptica*, have a very higher biomedical importance (7, 8). According to the previously published reports, *Bordetella* species have been mainly found as pathogens, but also various environmental samples; soil, water, and air are regarded as their habitats (7, 8). Recent findings suggest soil as a probable environmental origin of *Bordetella* species, including the animal-pathogenic lineages (7-9). The significant abundance of pathogenic *Bordetella* species in soils explains their wide distribution as well as frequent disease outbreaks that start without an obvious infectious source (9, 10).

*B. pertussis* is a strict human pathogen causing the respiratory tract infection called whooping cough (9). *B. parapertussis* consists of two lineages, one infecting human and the other infecting sheep (10). In contrast to these single host adapted pathogens, *B. bronchiseptica*: a close species to the two above species can cause a broad array of respiratory diseases (11, 12). *B. trematum* is a nonpathogenic, opportunistic organism whose sole source of isolation is thought to be open wounds of humans (13). In place, *B. trematum* causes ear and wound infections (14). A selective microbe-host association between *B. trematum* and *B. holmesii* species, and humans seems probable since these two *Bordetella* species have been exclusively detected as human pathogens (14). *B. holmesii* has been found repeatedly in blood and often in sputum of adolescents and is an emerging cause of septic arthritis (15-17). *B. avium*, a pathogen of birds, causes coryza or rhinotracheitis in poultry, but it has never been found in humans. *B. petrii*, causes sinusitis in immunocompromised adolescents, has been isolated from environmental samples and is capable of anaerobic growth (8, 18). *B. hinzii*, mainly colonizes the respiratory tract of poultry, has been also found as a chronic cholangitis infection agent in immunocompromised humans (19) and was recently reported as a causative agent of fatal septicemia (20). Since *B. hinzii* has been isolated from trachea and lungs of laboratory mice with respiratory infection and wild rodents, it is assumed that these animals may serve as reservoir for this species that could be transmitted to human or pets (19, 20). *B. hinzii* should be added to the list of emerging bacterial zoonotic agents in wild rodents that could be pathogenic for humans, especially immunocompromised patients (20, 21). *B. pseudohinzii*: a close species to *B. hinzii*, is also detected as a rodent-associated *Bordetella* species (19-22). *B. bronchialis*, *B. flabilis*, and *B. sputigena* (23, 24) are recently introduced and they have been isolated from human respiratory specimens. In contrast to other bordetellae, *B. trematum* (21) and *B. anisorpii* (22) are not associated with respiratory problems but are isolated from human wound infections.

Species delimitation seems to be difficult dealing with bordetellae. *B. hinzii* is highly difficult to become differentiated from *B. pseudohinzii* and even *B. avium* by routine phenotypic methods. Similarly, miss-identification is highly probable differentiating *B. parapertussis*, *B. pertussis* and *B. bronchiseptica* (24, 25). Sequence-based identification and phylogeny tend to be a promising approach to resolve the species boundaries (26).

Considering the increasing rate of the researches performed on various bacterial species to fulfill sequence-based identifications, the phylogenetic species boundaries have become faint. Thus, a single genomic locus may become exhausted from the needed signals to resolve very close species. Confirmation can be reliably achieved using
advanced genotypic and phylogenetic methods (24, 27), and the greater nucleotide variation of the conserved protein coding genes allows unequivocal identification of very close *Bordetella* species. Thus, in this study we performed a preliminary research on the applicability of *ompA* gene sequence, encoding a porin-like protein which has a critical role in pathogenesis, in phylogeny and identification of *Bordetella* species.

**Materials and Methods**

The 16S rRNA gene reference sequences and either the same gene sequences from isolates and uncultured materials obtained from Ribosomal Database Project (27, 28). Also, *ompA* gene sequences obtained from the nucleotide database of GenBank, NCBI. Thus, three different datasets (two datasets for 16S rRNA gene and a dataset for the nucleotide sequence of the coding gene for *ompA*) were prepared, separately.

The datasets were aligned with the multiple sequence alignment tool; Multiple sequence Alignment using Fast Fourier Transform (MAFFT), available at the European Bioinformatics Institute (EMBL-EBI), separately (29-35). Alignments were manually improved in MEGA v. 7.0.9 and Bioedit v. 7.0.5.3 packages (default settings) (36, 37). Maximum likelihood and Bayesian analyses were conducted using separated or concatenated datasets. The online tool Findmodel (http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html) was used to determine the best nucleotide substitution model for each partition. Bayesian inference (BI) analysis was conducted for each dataset, separately. Bayesian analyses were conducted with MrBayes v3.2.1 (38) executed on XSEDE (Extreme Science and Engineering Discovery Environment) through the CIPRES Science Gateway v. 3.3 (39) in two parallel runs, using the default settings but with the following modifications: general time reversible (GTR) model of DNA substitution as the best fit and a gamma distribution rate variation across sites (29). This model was chosen as the result from a pretest with MrModeltest v. 2.2 (40). After this was determined, the GTR + I + G model, as the best nucleotide substitution model, was used for the combined dataset, and a MCMC heated chain was set with a temperature value of 0.05. The number of chains, number of generations, and sample frequencies were set, respectively, at 4, 20000000 or 50000000, and 1000. Chain convergence was determined using Tracer v. 1.5 (http://tree.bio.ed.ac.uk/software/tracer/) to confirm sufficiently large ESS values (>200). The sampled trees were subsequently summarized after omitting the first 25% of trees as burn-in using the “sump” and “sumt” commands implemented in MrBayes (41, 42). The tree was visualized and edited using FigTree v. 1.4.2 (43, 32).

**Results**

Sequence dataset of the 16S rRNA gene which was provided by RDP database contained sequences of type strains, isolates and uncultured sequence data. The sequence data of this gene was screened and split into three separate alignments; type strains, isolates and uncultured sequence. Besides, sequence dataset of *ompA* gene was produced using the similarity search engines of BLAST program available at NCBI. The tree topology for the three separate alignments of the 16S rRNA gene was the same. Thus, the 16S rRNA gene alignments of the type strains and isolates were fused and used for further analyses (Fig. 1). Also, to infer the familial placement of the genus *Bordetella*, 16S rRNA gene sequences for the type strains of the genus *Bordetella* were analyzed in an alignment which contained the reference sequences for the genera of Alcaligenaceae and allied families.
Comparative Phylogeny of the Genus Bordetella ...


Badamchi A & Papizadeh M.
Figure 1. The Bayesian inference phylogeny of the members of the genus \textit{Bordetella} based on the 16S rRNA gene sequence data. Bayesian posterior probabilities above 0.75 resulting from 50,000,000 replicates are given at the nodes. The GenBank accessions are given after the species names. Species are differentiated with alternative colours (putative undescribed species are not highlighted). Type strains of the described \textit{Bordetella} species are shown in bold. The tree is rooted to \textit{Alcaligenes fæcalis}.
Figure 2.  The Bayesian inference phylogeny of the genus *Bordetella* based on the 16S rRNA gene sequence data of type strains. Bayesian posterior probabilities above 0.75 resulting from 20,000,000 replicates are given at the nodes. The GenBank accessions are given after the species names. The genera *Bordetella* and *Achromobacter* are differentiated with alternative colours. The tree is rooted to *Dexia gummosa*. 
Comparative Phylogeny of the Genus Bordetella …

Badamchi A & Papizadeh M.

Bordetella pertussis strain H788
Bordetella pertussis strain I498
Bordetella pertussis strain I669
Bordetella pertussis strain I707
Bordetella pertussis strain H800
Bordetella pertussis strain H812
Bordetella pertussis strain I228
Bordetella pertussis strain I373
Bordetella pertussis strain I752
Bordetella pertussis strain H320
Bordetella pertussis strain H834
Bordetella pertussis strain H710
Bordetella pertussis strain H851
Bordetella pertussis strain H915
Bordetella pertussis strain I975
Bordetella pertussis strain J022
Bordetella pertussis strain I344
Bordetella pertussis strain C505
Bordetella pertussis strain C742
Bordetella pertussis strain D175
Bordetella pertussis strain D321
Bordetella pertussis strain H346
Bordetella pertussis strain H437
Bordetella pertussis strain H775
Bordetella pertussis strain H779
Bordetella pertussis strain H787
Bordetella pertussis strain H864
Bordetella pertussis strain I110
Bordetella pertussis strain I127
Bordetella pertussis strain I136
Bordetella pertussis strain I331
Bordetella pertussis strain I375
Bordetella pertussis strain I380
Bordetella pertussis strain I386
Bordetella pertussis strain I728
Bordetella pertussis strain I755
Bordetella pertussis strain I959
Bordetella pertussis strain I965
Bordetella pertussis strain I968
Bordetella pertussis strain J023
Bordetella pertussis strain J076
Bordetella pertussis strain H681
Bordetella pertussis strain I751
Figure 3. The Bayesian inference phylogeny of the members of the genus *Bordetella* based on the sequence data of the coding gene for *ompA*. Bayesian posterior probabilities above 0.75 resulting from 20,000,000 replicates are given at the nodes. The GenBank accessions are given after the species names. Species are differentiated with alternative colours. The tree is rooted to *Achromobacter insolitus* (DSM 23807).
Table 1. Genomics data of some *Bordetella* species available in the genome database of GenBank, NCBI.

<table>
<thead>
<tr>
<th>Species</th>
<th>Genome assemblies</th>
<th>Median total length (Mb)</th>
<th>Median protein count</th>
<th>Median GC content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. bronchiseptica</em></td>
<td>68</td>
<td>5.19</td>
<td>4759</td>
<td>68.2</td>
</tr>
<tr>
<td><em>B. parapertussis</em></td>
<td>4</td>
<td>4.78</td>
<td>4162</td>
<td>68.1</td>
</tr>
<tr>
<td><em>B. pertussis</em></td>
<td>613</td>
<td>4.05</td>
<td>3576</td>
<td>67.7</td>
</tr>
<tr>
<td><em>B. holmesii</em></td>
<td>21</td>
<td>3.61</td>
<td>3139</td>
<td>62.7</td>
</tr>
<tr>
<td><em>B. hinzii</em></td>
<td>10</td>
<td>4.89</td>
<td>4456</td>
<td>67</td>
</tr>
<tr>
<td><em>B. petrii</em></td>
<td>3</td>
<td>5.04</td>
<td>4718</td>
<td>65.5</td>
</tr>
<tr>
<td><em>B. avium</em></td>
<td>2</td>
<td>3.71</td>
<td>3262</td>
<td>61.6</td>
</tr>
<tr>
<td><em>B. pseudohinzii</em></td>
<td>4</td>
<td>4.53</td>
<td>4124</td>
<td>66.6</td>
</tr>
<tr>
<td><em>B. trematum</em></td>
<td>5</td>
<td>4.44</td>
<td>3985</td>
<td>65.7</td>
</tr>
<tr>
<td><em>B. anisorpii</em></td>
<td>2</td>
<td>6.17</td>
<td>5357</td>
<td>66.8</td>
</tr>
<tr>
<td><em>B. flabilis</em></td>
<td>2</td>
<td>5.95</td>
<td>5238</td>
<td>65.9</td>
</tr>
<tr>
<td><em>B. bronchialis</em></td>
<td>2</td>
<td>5.92</td>
<td>5144</td>
<td>67.3</td>
</tr>
</tbody>
</table>

An intraspecies diversity can also be observed in *B. petrii* clade (Fig. 1).

Phylogenies performed on the coding gene for *ompA* confirmed the efficient variability of the nucleotide sequence of this gene which resolve all *Bordetella* species as very well supported clades (Fig. 3). Moreover, the tree topology of *ompA* based phylogenies was conforming to that of 16S rRNA gene.

Abundance of the sequence data of these two genes of *Bordetella* species in the nucleotide database of GenBank, NCBI is not comparable. In fact, there were only 83 (62 sequences from *B. pertussis* and 21 sequences from other *Bordetella* species) nucleotide sequences of the coding gene for *ompA* belonging to *Bordetella* species. In comparison, there were 247 16S rRNA sequences from *Bordetella* species which were analyzed in our phylogenies (Fig. 1).

16S rRNA based phylogeny showed that there are still some clades in *Bordetella* which seem to be putative undescribed species. However, *ompA* didn’t show further data on the diversity and boundaries of the genus which is highly associated with the under-sampling of the nucleotide sequences of this gene (Fig. 3).

Discussion

Analyzing the 16S rRNA gene alignment, it was found that this gene, as the main gene in phylogeny purposes in prokaryotes, has some limitations to resolve *Bordetella* species. This weak point of the 16S rRNA gene is very well highlighted in figure 1, where two of the three most important medical species: *B. bronchiseptica* and *B. parapertussis* were not resolved.

Our results show that the *Bordetella* species have been mostly detected in soil, water, sediment, and even associated to some plants, worldwide. Further, considering Fig. 1, it is shown that human/animal-associated *Bordetella* species scatter in the phylogenetic tree of the genus and it is contrary to the results of Soumana et al. (44).

Furthermore, phylogenies conducted in this study indicated that the *Bordetella* species with in some basal positions to the rest of the genus (*B. bronchialis*, *B. flabilis*, *B. sputigena*) have been exclusively detected in human respiratory specimens (24). Thus, the conclusion that the basal clades harbor species with environmental origins is still discussed and it may be in contrary to conclusion of Soumana et al. (2017) (44). According to the data summarized in table 1, the above mentioned species have larger genomes comparing to the rest of the genus. Of course,
phylogenies conducted in this study are conforming to those of Vandamme et al. (2015) as these three species have a different node from the other Bordetella species (23). Thus, more taxonomic revisions seems plausible. According to the recent 16S rRNA-based phylogenies, it was assumed that Bordetella species with environmental origins tend to have basal placements in comparison to human/animal-associated species (44), but the gigantic phylogenies performed here and the results of Vandamme et al. (2015) does not show such a relationship between the origin of the Bordetella species/isolates and their evolutionary placements (23). Also, our phylogenies showed that there are a considerably higher genetic diversity in the basal taxa of the phylogenetic tree of Bordetella which is conforming to the results of Soumana et al. (44). However, considering the documented genome decay rates in Bordetella species, an evolutionary link between species with a free-living environmental lifestyle and the species with a host-restricted obligately pathogenic lifestyle is probable.

**Conclusion**

As a conclusion, considering the analyses performed on the nucleotide sequences of the coding gene for *ompA* a higher resolution achieved for Bordetella species. Also, due to the same topologies observed for 16S rRNA and *ompA* genes it is concluded that using coding genes; likely *ompA*, can result more resolutions in Bordetella phylogenies which differentiate very close species unequivocally.

**Acknowledgements**

This research was supported by Pasteur Institute of Iran (IPI), Tehran, Iran.

**Conflict of interest**

None declared conflicts of interest.

**References**

9. Lehmann PF, PR Murray, EJ Baron, et al,


Comparative Phylogeny of the Genus Bordetella ...

Badamchi A & Papizadeh M.


