A Simplified Sequence-Based Identification Scheme for Bordetella Reveals Several Putative Novel Species

Theodore Spilker, Amy L. Leber, Mario J. Marcon, Duane W. Newton, Rebecca Darrah, Peter Vandamme, and John J. LiPuma

Department of Pediatrics and Communicable Diseases, University of Michigan Medical School, Ann Arbor, Michigan, USA; Department of Laboratory Medicine, Nationwide Children's Hospital, Columbus, Ohio, USA; Department of Pathology, University of Michigan Medical School, Ann Arbor, Michigan, USA; Frances Payne Bolton School of Nursing, Case Western Reserve University, Cleveland, OH, USA; and Laboratory of Microbiology, Ghent University, Ghent, Belgium

Running title: nrdA-based identification of Bordetella species

Correspondent footnote: jlipuma@umich.edu
(Abstract)
The differentiation of *Bordetella* species, particularly those causing human infection, is problematic. We found that sequence analysis of an internal fragment of *nrdA* allowed differentiation of the currently named *Bordetella* species. Analysis of 107 “*Bordetella*” isolates recovered almost exclusively from human respiratory tract specimens identified several putative novel species.
The genus *Bordetella* currently consists of eight species: *Bordetella avium*, *Bordetella bronchiseptica*, *Bordetella hinzii*, *Bordetella holmesii*, *Bordetella parapertussis*, *Bordetella pertussis*, *Bordetella petrii* and *Bordetella trematum*. Another species, "*Bordetella ansorpii*", has been proposed, but not validly named, and is represented by only a couple reported isolates (1). Historically, the need to reliably detect and differentiate *Bordetella* species in clinical samples has been limited by the predominant role that *B. pertussis* was believed to play in human infection. However, other *Bordetella* species appear to be involved in human infection more commonly than previously appreciated. *B. parapertussis*, in particular, likely accounts for a considerable minority of cases of whooping cough (pertussis) (2) and, recently, *B. holmesii* was identified in a large community outbreak of a pertussis-like illness in Ohio (3). Currently, the most common diagnostic testing involves various real-time polymerase chain reaction (RT-PCR) assays targeting pertussis toxin or toxin promoter genes and/or variable number insertion elements to detect and differentiate *Bordetella* species directly in clinical specimens. These assays, together with combinations of growth properties, biochemical tests, serotyping, and sequencing of a portion of the 16S rRNA gene, are used to differentiate *Bordetella* species recovered in culture. However, these methods are cumbersome and specific interpretive criteria for species differentiation are lacking (4). We recently showed that sequence analysis of *nrdA*, which encodes the ribonucleoside-diphosphate reductase alpha chain, can reliably differentiate species within the genus *Achromobacter* (5, 6). Given the phylogenetic relatedness of *Bordetella* and *Achromobacter*, we sought to determine if sequence analysis of this locus would similarly allow reliable differentiation of cultured isolates of...
Bordetella species, thereby providing the basis for further analysis of this target for direct detection of Bordetella in clinical specimens.

The 165 strains analyzed in this study are listed in the supplemental material (Table S1). A total of 107 isolates recovered from culture of biologic specimens primarily from persons with and without cystic fibrosis (CF) were referred to the Burkholderia cepacia Research Laboratory and Repository (University of Michigan) from 35 clinical microbiology laboratories in 23 states in the USA between 2001 and 2013. DNA preparations from 30 Bordetella strains cultured primarily from human nasopharyngeal swabs were obtained from Nationwide Children’s Hospital (Columbus, Ohio). These 30 strains had been identified as B pertussis, B. parapertussis or B holmesii by traditional and molecular methods (3). Sixteen type and taxonomic reference strains were obtained from the BCCM/LMG culture collection (Ghent, Belgium). DNA sequences from the 12 remaining Bordetella strains were obtained from whole genome sequences available at the National Center for Biotechnology Information, USA (NCBI; http://www.ncbi.nlm.nih.gov/genome/browse/).

All bacterial cultures were incubated aerobically at 32°C for 24 h on Mueller-Hinton agar. DNA was prepared from single colonies as described previously (5). PCR primers P1 and P2, with the following 5’ → 3’ nucleotide sequences, were manually designed to target nrdA: P1, GCCCGACCTGCACGAC; P2, CCGGASAGRTTCGATTTGACGTAC. DNA amplification and sequencing were performed as previously described with the following modification (5). After an initial denaturation for 2 min at 95°C, 30 cycles were completed with each consisting of 30 s at 94°C, 30 s at 58°C, and 60 s at 72°C. A final extension of 5 min at 72°C was applied with an infinite hold at 8°C. The bacterial 16S
rRNA gene was amplified by PCR as previously described (7). DNA sequencing and editing of sequence chromatograms was performed as described (5). nrdA sequences were trimmed to 765 bp and aligned using MegAlign (DNASTar, Madison, WI, USA) with Clustal W to generate dendrograms with 1,000 bootstrap replications using the MegAlign default parameters. For validation of distinctness of sequence similarity clusters, the $k$ parameter was calculated as previously described (8).

All nrdA sequences generated in this study can be found at the PubMLST site at http://pubmlst.org/bordetella (9). The 16S rRNA gene sequences generated in this study were deposited in GenBank (http://www.ncbi.nlm.nih.gov/genbank) and are assigned the following accession numbers: EU082134, EU082135, EU082146, EU082151, EU082156, EU082157, EU082159, EU082162, and KF601902 through KF601915.

A dendrogram based on the 765 bp fragment of the nrdA gene from 12 Bordetella strains for which whole genome sequences were available from NCBI, together with sequences from 16 taxonomic reference strains and 30 Bordetella strains previously identified to the species level (3), separated strains into clusters representing each of the named species in the genus (Fig. 1 and Fig. S1). As expected, B. pertussis, B. parapertussis and B. bronchiseptica, showed greater similarity at this locus relative to the other species in the genus. Two distinct clusters of B. bronchiseptica were apparent, in agreement with the separation of this species into two groups (“complexes I and IV”) described by Diavatopoulos and colleagues (10), based on sequence analysis of seven housekeeping genes.
When the 107 clinical isolates were added to the 58 reference strains described above, the *nrdA* sequence-based dendrogram again showed clustering corresponding to the named species in the genus (Fig. S2). However, several strains grouped into new clusters apart from those representing the named *Bordetella* species. Several individual strains did not cluster with any others, but rather occupied distinct branches in the tree. Nine clusters were each comprised of at least two strains with different *nrdA* sequences, allowing for calculation of a *k* parameter to discern how distinct these were from each other and from the nine named *Bordetella* species. The *k* parameters for these nine clusters ranged from 3.88 to 167.43 (Table 1), indicating that each of these indeed represented a distinct genogroup, designated *Bordetella* genogroups 1 through 9.

Another cluster consisted of two strains with the same *nrdA* sequence, precluding calculation of a *k* parameter. Nevertheless, for the sake of description, this cluster was designated genogroup 10. The six strains occupying solitary branches were designated genogroups 11-16.

A dendrogram based on 16S rRNA gene sequences was constructed to include one strain from each named *Bordetella* species, including *B. bronchiseptica* complex I and *B. bronchiseptica* complex IV, as well as one representative from each of the 16 novel genogroups (Fig. S3). As expected, *B. parapertussis*, *B. bronchiseptica* complex I, and *B. bronchiseptica* complex IV could not be distinguished from each other at this locus. Of note, *nrdA* genogroup 6 was also indistinguishable from *B. parapertussis* and *B. bronchiseptica* by 16S rRNA gene sequence analysis.

Among the several species of human and animal pathogens in the genus *Bordetella*, *B. pertussis* continues to garner the most attention in human health, particularly with the
increased incidence of pertussis in the United States in recent years despite the routine
use of an effective vaccine (11). B. parapertussis and B. holmesii can also cause a
pertussis-like illness that may be more common than generally appreciated (2).
Recently, Rodgers and colleagues (3) reported a community outbreak of a pertussis-like
illness in Ohio involving 918 persons. Among the NP swab samples available from 298
persons, Bordetella species were detected in 164 based on a positive IS481 PCR
assay. Using multiple PCR assays, 68% of these samples were confirmed as B.
pertussis, 29% as B. holmesii, and 2% were confirmed as containing both species.
The assessment of the prevalence of these species in human respiratory illness, as
well as a better understanding of the natural history of Bordetella infection and analyses
of pertussis vaccine efficacy are predicated on reliable methods to detect and
differentiate these species; however, this can be challenging. Currently, an RT-PCR
assay targeting IS481 is most often used to identify B. pertussis, which contains 50-238
copies of this insertion element, in clinical specimens (12, 13). However, IS481 or IS481-
like sequences can be found in B. holmesii and in some strains of B. bronchiseptica
(14). Thus, without further differentiation, IS481-positive samples may be falsely
reported as B pertussis. Additional PCR assays to detect and differentiate Bordetella
species include those targeting the insertion elements hIS1001 (found in B. holmesii
only), pIS1001 (found in B. parapertussis and infrequently in B. bronchiseptica) (15).
These assays may be combined with PCR assays targeting the toxin gene ptxS1
(detected in B. pertussis, B. parapertussis and B. bronchiseptica) and/or intact ptxA-Pr
sequences (detected in B. pertussis). Mutated ptxA-Pr sequences may be found in B.
parapertussis and B. bronchiseptica. Despite the use of these assays, the molecular
detection and differentiation of Bordetella species remains problematic (4, 16).

To develop a reliable streamlined assay to differentiate Bordetella species recovered
in culture, we drew upon our recent experience (5, 6) using analysis of a single genomic
locus to identify species in the genus Achromobacter, which is phylogenetically closely
related to Bordetella. In that work, we found that DNA sequence analysis of a 765 bp
segment of nrdA enabled differentiation of the seven named Achromobacter species
that was comparable to the separation of these species provided by multilocus
sequence analysis. Further, in an analysis of 147 strains, we identified 14 novel
Achromobacter genogroups, several of which were found to cause infection in persons
with CF (6). In subsequent and on-going work, comprehensive taxonomic assessment
has confirmed that these groups represent novel Achromobacter species. Eight of
these species now have been formally described (17, 18).

In the present study, we found that nrdA sequence analysis differentiated the
currently named Bordetella species (Fig. 1). This includes separation of B. pertussis, B.
parapertussis, and the two B. bronchiseptica subpopulations (complexes I and IV)
described by Diavatopoulos and colleagues (10). All B. pertussis strains were clonal at
this locus, even though these were from geographically diverse locations (USA,
Netherlands, China, and Japan), with some being initially recovered more than 60 years
ago. All B. parapertussis also had identical sequences at this locus, as did the seven
available B. holmesii isolates. In contrast, 12 alleles were found among the 63 B.
bronchiseptica isolates; 3 alleles were found among the 11 B. hinzii, and 3 were noted
in the four B. trematum included in the study.
In extending our analysis to a larger collection of *Bordetella* isolates cultured from human respiratory specimens, most of which were from persons with CF, we noted that several groups or individual strains appeared to cluster apart from the known *Bordetella* species on the *nrdA*-based dendrogram. To assess the distinctness of these groups, we calculated the $k$ parameter, which is the ratio of the mean inter-group sequence divergence to the mean intra-group sequence divergence (8). A $k$ parameter value greater than 2 indicates that a group is distinct from its closest neighbor. We found that genogroups 1 through 9, each of which included strains with two or more different *nrdA* sequences, would all be considered distinct based on their respective $k$ parameters, indicating that each most likely represents a novel species. Analysis of a greater number of strains that cluster with isolates in genogroups 10 through 16 will be required to determine if these too are sufficiently genetically distinct to represent novel taxa.

In summary, we have shown that *nrdA* sequence analysis allows for the reliable differentiation of the currently named *Bordetella* species recovered in culture. In addition, we have identified several distinct genogroups among "*Bordetella*-like" isolates recovered from human respiratory specimens. Analysis of the 16S rRNA gene from representatives of each of these novel genogroups (Fig S3) suggests that some (genogroups 3, 8, 9, 10, 11, 14, and 15) may represent a novel genus, closely related to *Achromobacter* and *Bordetella*. Clarification of the taxonomy of these genogroups will provide the basis for further studies to assess the utility of *nrdA* amplification and sequence analysis for the direct detection and differentiation of *Bordetella* species in respiratory specimens.
ACKNOWLEDGMENTS

This work was supported by the Cystic Fibrosis Foundation.
REFERENCES


FIG. 1. *nrdA*-based dendrogram containing two (if available) representative strains of the nine named species in the genus *Bordetella*. *B. bronchiseptica* complex I and *B. bronchiseptica* complex IV are represented (10). The tree is rooted with *Achromobacter xylosoxidans*. 
<table>
<thead>
<tr>
<th>Reference group (no. strains available)</th>
<th>Closest neighbor</th>
<th>intragroup divergence</th>
<th>intergroup divergence</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B.$ genogroup 1 (5)</td>
<td>$B.$ genogroup 5</td>
<td>0.46</td>
<td>8.64</td>
<td>18.78</td>
</tr>
<tr>
<td>$B.$ genogroup 2 (7)</td>
<td>$B.$ petrii</td>
<td>0.06</td>
<td>5.93</td>
<td>98.83</td>
</tr>
<tr>
<td>$B.$ genogroup 3 (8)</td>
<td>$B.$ genogroup 10</td>
<td>0.07</td>
<td>9.44</td>
<td>134.86</td>
</tr>
<tr>
<td>$B.$ genogroup 4 (3)</td>
<td>$B.$ genogroup 7</td>
<td>0.07</td>
<td>11.72</td>
<td>167.43</td>
</tr>
<tr>
<td>$B.$ genogroup 5 (2)</td>
<td>$B.$ genogroup 1</td>
<td>0.10</td>
<td>8.64</td>
<td>86.40</td>
</tr>
<tr>
<td>$B.$ genogroup 6 (8)</td>
<td>$B.$ bronchiseptica I</td>
<td>0.29</td>
<td>2.63</td>
<td>9.07</td>
</tr>
<tr>
<td>$B.$ genogroup 7 (3)</td>
<td>$B.$ genogroup 4</td>
<td>0.07</td>
<td>11.67</td>
<td>166.71</td>
</tr>
<tr>
<td>$B.$ genogroup 8 (2)</td>
<td>$B.$ genogroup 14</td>
<td>1.30</td>
<td>5.05</td>
<td>3.88</td>
</tr>
<tr>
<td>$B.$ genogroup 9 (2)</td>
<td>$B.$ genogroup 3</td>
<td>0.30</td>
<td>10.24</td>
<td>34.13</td>
</tr>
</tbody>
</table>

TABLE 1. $k$ parameters for novel *Bordetella* genogroups.
FIG. 1.

Nucleotide Substitution per 100 residues

B. parapertussis 31129
B. parapertussis 12822
B. bronchiseptica AU4851
B. bronchiseptica LMG 3531
B. pertussis LMG 14455
B. pertussis Tohama I
B. bronchiseptica LMG 3521
B. bronchiseptica AU8057
B. holmesii 29465
B. holmesii LMG 15945

Complex I

B. hinzii LMG 13500
B. hinzii AU11496
B. trematum LMG 13506
B. trematum 32223
B. avium197N

Complex IV

B. petrii DSM 12804
“B. ansorpii” R-22284
Achromobacter xylosoxidans LMG 1863

11.7

Nucleotide Substitution per 100 residues