Simple Algorithm for Identification of Bordetella pertussis Pertactin Gene Variants

Gaëtan Muyldermans,1 Denis Piérard,1* Nathalie Hoebrekx,1 Reza Advani,2 Shirley Van Amersfoorth,3 Iris De Schutter,4 Oriane Soetens,1 Leo Eeckhout,1 Anne Malfroot,4 and Sabine Lauwers1

Belgian Reference Laboratory for Pertussis, Department of Microbiology,1 and Department of Pediatrics,4 Academisch Ziekenhuis Vrije Universiteit Brussel, Brussels, Belgium; Swedish Institute for Infectious Disease Control, Solna, Sweden2; and Research Laboratory for Infectious Diseases, National Institute for Public Health and Environment (RIVM), Bilthoven, The Netherlands3

Received 18 July 2003/Returned for modification 5 September 2003/Accepted 12 January 2004

Studies performed in several countries have demonstrated the recent emergence and subsequent dominance of circulating Bordetella pertussis strains harboring pertactin and pertussis toxin variants not included in pertussis vaccines. Determination of the pertactin gene variants is commonly performed using a time-consuming and expensive sequence analysis. We developed a simple and reliable pertactin typing algorithm suitable for large-scale screening. The assay correctly identified all pertactin alleles in representative strains. The typing of 231 clinical strains of B. pertussis routinely isolated in Belgium showed that this algorithm was adequate to identify less-frequent prn types like prn9 and prn11.

Pertussis (whooping cough), mainly caused by Bordetella pertussis, is still responsible for high morbidity and mortality among children in many parts of the world. The re-emergence of pertussis in several countries with high vaccine coverage (1, 3, 6) has led to renewed interest in whooping cough and in pertussis vaccine efficacy. Vaccine-induced immunity may select among circulating strains for antigenic variants that are divergent from those of the vaccine strains (2, 5, 10, 12, 13). In a recent multicollateral sequence typing study based on 15 genes coding for surface proteins, polymorphism was observed in the genes of pertactin (prn), pertussis toxin subunits S1 and S3 (ptsS1 and ptsS3), and the tracheal colonizing factor (tcf4) (14).

Pertactin, a 69-kDa outer membrane protein, is a major virulence factor of B. pertussis that mediates adherence to host cells through an arginine-glycine-aspartic acid (RGD) sequence (8). It is a component of many acellular subunit vaccines. On the basis of polymorphism that occurs mainly in the repetitive region (11), 11 prn alleles showing a variable number of GGXXP repeats (Fig. 1A) have been identified (EMBL GenBank database). As demonstrated in isolates collected in The Netherlands and Finland (13), all isolates from the prevaccination period harbored the vaccine pertactin type prn1. However, these types were replaced during the 1990s by the nonvaccine prn types prn2 (72 and 27%, respectively, in these two countries), prn3 (12 and 63%, respectively), and prn4 (0 and 9%, respectively). For the same period, high prevalences of nonvaccine strains were detected as well in other countries such as Italy (6% prn1, 41% prn2, 51% prn3, and 2% prn5), the United States (30% prn1 and 70% prn2), and the United Kingdom (53% non-prn5) (2, 5, 10). The allelic variation in prn types prn1 to prn4, which represent more than 98% of tested clinical isolates, is restricted to region 1. Typing of these alleles necessitates, therefore, the characterization of this region. However, some less frequent prn types (prn5 to prn11) showing variation in other regions have been identified. The determination of the single nucleotide polymorphism T1595G, present in prn6 to prn8 and prn10, provides a way to further discriminate these rare prn types against the predominant prn types prn1 to prn4.

After the introduction of the whole-cell vaccine in the 1950s, a progressive shift toward alleles not found in the vaccine strains has been observed. Since the currently circulating isolates differ from the vaccine strains (2, 4, 5, 10, 12, 13, 15), it is important to monitor the corresponding variations, particularly in view of the recent introduction of acellular vaccines in most national immunization programs. The composition of these vaccines is not uniform and differs in the number of antigens, which varies from 1 to 5, and in the antigenic subtypes included. The number of antigens or the inclusion of pertactin (only in vaccines with three or more components) may be responsible for differences in efficacy (7).

Monitoring of clinical isolates by simpler alternatives to sequence analysis improves the turnaround time and expands the analytical capacity to additional strains. So far, the reference method for the determination of the pertactin type has been the sequencing of the prn gene (11), a relatively time-consuming and expensive method. Recently, Mäkinen et al. (9) developed a rapid identification method based on real-time PCR in combination with gel electrophoresis. Although the speed and simplicity of this approach make it an advantageous alternative to conventional sequencing of the prn gene, this method needs an expensive real-time apparatus and does not differentiate new types such as prn6 to prn11.

We developed a simple method for discriminating the most
FIG. 1. (A) Polymorphism in the prn gene region 1. The 11 known sequences are aligned. Dots represent identical bases; dashes represent deleted bases. The nucleotide numbers correspond to those of prn1 sequence accession number AJ011091 (GenBank database [http://www.ncbi.nlm.nih.gov/entrez]). (B) Primers prnC and prnE were designed to recognize a region of two constitutive GGAVP repeats and the overlap between GGAVP and GGFGP, respectively. They were used in two separate reactions with the prnD primer that hybridizes with a region downstream of region 1, as shown in panel C. (C) The amplification of isolates with primer pairs prnC-prnD and prnE-prnD allowed us to determine the total number of repeating units (GGAVP and GGFGP) and the number of GGFGP repeating units, respectively (as at right). n. a., no amplification.
frequent prn2 and prn3 types from the other less frequently occurring variants and to differentiate prn2 and prn3 from each other. Further identification of the few remaining isolates can then be performed by either sequencing or real-time PCR. A complete algorithm is proposed and successfully applied to 231 B. pertussis isolates, representing almost all the strains isolated in Belgium from 1987 to 2001.

MATERIALS AND METHODS

Strains. The analysis was initially performed on a set of strains representing the different prn types kindly provided by J. Mertsola (University of Turku, Turku, Finland) and F. Mooi (National Institute for Public Health and the Environment, Bilthoven, The Netherlands). In addition, 231 clinical B. pertussis isolates collected from 1987 through 2001 in our laboratory were used for this study.

Bacteria were grown on Regan-Lowe charcoal agar (Charcoal Agar; Oxoid Ltd., Basingstoke, England) containing 10% horse blood and cephalin (Bordetella Selective Supplement), incubated during 4 days at 35°C. For PCR, one isolated colony was resuspended in TE (10 mM Tris-HCl [pH 8], 1 mM EDTA) at 4°C for 2 min. The amplification product was further analyzed on 3% LSI MP agarose (Life Science International, Zellik, Belgium) gel, and the length of the DNA band was compared with those of control samples from prn1 to prn6.

Real-time PCR for the identification of the T1595G point mutation. The real-time PCR method was adapted from that of Mäkinen et al. (8) for SYBR Green I DNA detection. Briefly, two amplifications were performed for the determination of the polymorphism at position 1595. Primer pairs QJF3-QJR1 and QJF4-QJR1 preferentially amplify (i) prn1 to prn3, prn9, and prn11 and (ii) prn6 to prn8 and prn10, respectively. The PCR mixture was optimized for an 1-color apparatus (Bio-Rad, Hercules, Calif.) using the qPCR Core Kit for SYBR Green I (Eurogentec, Seraing, Belgium) with the addition of 0.01 μM fluorescein and 1 μl of boiled sample. After an initial denaturation step at 94°C for 10 min, 50 cycles of denaturation at 94°C for 15 s, annealing at 62°C for 15 s, and extension at 72°C for 30 s were performed. Measurement of fluorescence at 485 nm at the end of each extension step allows the detection of the SYBR Green I bound to the amplified double-stranded DNA. The increase in fluorescence signal during the PCR process correlates with PCR product accumulation. The parameter used to type the single nucleotide polymorphism was the cycle threshold (Ct). This is the cycle number at which the reaction begins to be exponential. It is determined by the intersection between the amplification curve and the Ct level, calculated as the 10-fold standard deviation of the observed fluorescence signal between cycles 5 and 20. QJF3 and QJF4 (Table 1) primers were designed to preferentially anneal and extend the alleles with T1595T and T1595G mutations, respectively. Using each of these primers in combination with the antisense QJR1 primer results in a difference in amplification efficiency. The PCR with the lowest Ct was considered to contain the primers with the most specific binding. After amplification, melting curve analysis of the PCR product was used to differentiate between specific and nonspecific amplification products and thereby confirming the previous amplification result. These curves were obtained by slowly changing the temperature of the reaction solution from 55 to 95°C while continuously collecting fluorescence data. This increase in temperature induces the PCR products to denature, which is accompanied by a decrease in the fluorescence from solutions containing the SYBR Green I dye. To improve the visualization of the melting temperature (Tm) melting peaks were derived from the initial melting curves (relative fluorescence units [RFU] versus temperature [°C]) by plotting the negative derivative of fluorescence over temperature versus temperature [−d(RFU)/dT versus °C].

Sequencing of the pertactin gene. Sequencing of the prn gene from isolates Bord4, Bord49, and Bord68 was performed as described by Mooi et al. (13).

Results

Table 1. Primers used in this studya

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’ to 3’)</th>
<th>Position</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>prnC</td>
<td>GTCCGCTTCGCGCGCGGTG</td>
<td>827–844</td>
<td>This study</td>
</tr>
<tr>
<td>prnD</td>
<td>GCTCCACGGTCGAGCGCGC</td>
<td>946–929</td>
<td>This study</td>
</tr>
<tr>
<td>prnE</td>
<td>TGGGTTCCCCGCGCGCTC</td>
<td>858–876</td>
<td>This study</td>
</tr>
<tr>
<td>QJF3</td>
<td>GCTGGTGCAGACGGCCAGT</td>
<td>1605–1622</td>
<td>This study</td>
</tr>
<tr>
<td>QJF4</td>
<td>GCTGGTGCAGACGGCCAGG</td>
<td>1605–1622</td>
<td>This study</td>
</tr>
<tr>
<td>QJR1</td>
<td>CGGATATCGACCTTGCC</td>
<td>1676–1690</td>
<td>9</td>
</tr>
</tbody>
</table>

a The position numbers indicate the position of bases relative to the first start codon of prn1 (accession number AJ011091). The differences between the primers prnC and prnE as well as those between QJF3 and QJF4 are shown in boldface type.

Sequence-specific PCR for typing the repetitive region 1. Two amplifications were performed for the determination of the polymorphism of the prn region 1 using either the prnC-prnD or prnE-prnD primer pair (Fig. 1B and C). The PCR method was optimized for the following variables: Mg concentration, annealing temperature, standard or hot-start PCR, number of cycles, and input concentration.

The 50-μl reaction mixture contained 1× Reaction Buffer II (Applied Biosystems, Foster City, Calif.), 1.5 mM MgCl2, a 200 μM concentration of each deoxynucleoside triphosphate, a 1 μM concentration of each primer, and 1 U of AmpliTaq Gold DNA polymerase (Applied Biosystems). Amplification was performed in a DNA Thermal Cycler 480 (Applied Biosystems) using the following program: 94°C for 10 min and then 40 cycles of 94°C for 1 min, 66°C for 1 min 30 s, and 72°C for 2 min.

The amplification product was further analyzed on 3% LSI MP agarose (Life Science International, Zellik, Belgium) gel, and the length of the DNA band was compared with those of control samples from prn1 to prn6. FIG. 2. Ethidium bromide-stained 3% LSI MP agarose (Life Science International) gel containing B. pertussis DNA amplified with primer pairs prnC-prnD (A) and prnE-prnD (B) from strains with known prn types.
This method allowed us to discriminate the 11 known 
prn types differing from the more-frequent 
prn types prn1 to prn5, prn9, and 
prn11 contain a T at position 1595, these type strains are 
preferentially recognized by the QJF3 primer, which contains a 
T at the 3' end. Preferential amplification is obtained with this 
sense primer in combination with the antisense QJR1 primer 
compared to the QJF4-QJR1 pair. On the other hand, as prn6 
to prn8 and prn10 contain a G at position 1595, these alleles are 
preferentially amplified by the QJF4-QJR1 primer pair. In Fig. 
4A a comparison of the amplification profiles from prn1 and 
prn7 are shown for the real-time amplification with the primer 
pairs QJF3-QJR1 and QJF4-QJR1. The lower value of the 
intersection between the curve and the Ct level with primer 
pair QJF3-QJR1 (30.7) compared to that for primer pair 
QJF4-QJR1 (37.4) confirms that T1595T is present in the 
prn1 allele. In contrast, the Ct for the prn7 using the primer 
pair QJF4-QJR1 is 15 cycles lower than that obtained with QJF3- 
QJR1. The melting curve analysis confirmed this interpreta-
tion, as a specific melting peak was observed for the prn1 strain 
amplified with the QJF3-QJR1 primer pair at 88.0°C and not for 
the QJF4-QJR1 primer pair (Fig. 4B). For the prn7 strain, 
a specific melting peak (88.5°C) was observed only with the 
QJF4-QJR1 amplification. No specific amplification could be 
detected for the negative control and the melting curve analysis 
did not show any peak (data not shown).

Table 2 summarizes the Ct for the representative strains and 
their T_m's. The mean T_m was 88.0°C (range, 87.5 to 88.5°C).

Use of the prn typing algorithm for typing clinical isolates. 
A total of 231 clinical isolates collected in Belgium from 1987 
through 2001 were typed using our algorithm (Fig. 3): 22 
(9.5%), 140 (60.6%), and 66 (28.6%) contained prn1, prn2, and 
prn3, respectively. The sequence-specific PCR allowed us to 
discriminate three isolates (Bord4, Bord49, and Bord68) with 
prn types differing from the more-frequent prn types prn1 to 
prn3. Since Bord4 and Bord68 contained a 149-bp fragment in 
the amplification reaction mixture with primer pair prnC-prnD 
(Fig. 5A), prn genes from both isolates were classified as prn9. 
The presence of the 134-bp fragment in the amplification re-
action mixture using the primer pair prnE-prnD (Fig. 5B) and 
the presence of the T1595T mutation confirmed this result 
(Table 2). On basis of the presence of the 134- and 89-bp 
fragments for the amplification reaction mixture using primer 
pairs prnC-prnD and prnE-prnD, respectively (Fig. 5), the prn 
gene from isolate Bord49 was classified as a prn10 or prn11. 
The presence of the T1595T mutation as shown by the real-
time PCR (Table 2) allowed us to further characterize this 
gene as a prn11. The type of these three rare prn variants was 
subsequently confirmed by sequence analysis (GenBank acces-
son numbers AY382174, AY382175, and AY382176).

DISCUSSION

Sequencing the prn gene in regions 1 and 2 is probably the 
most accurate method for typing the prn gene. It was previ-
ously demonstrated in numerous studies that prn1, prn2, and 
prn3 types predominate in different parts of the world (2, 4, 5, 
10, 12, 13, 15). Since these types differ only in region 1, it was 
proposed that region 1 in all isolates should be sequenced and 
when a novel variant of region 1 was found, that the complete 
prn gene should be sequenced (11). However, this method is 
time-consuming and expensive.
In this study, a simple algorithm was developed for typing the prn gene from B. pertussis. The method proved to be accurate and has a short turnaround time of 1 day. It allows the differentiation of the common prn types prn1 to prn3 from each other. Although prn6 and prn7 could not be differentiated, the method allows the identification of less-frequent prn types. Furthermore, it is convenient for large-scale screening while rapidly providing an epidemiologic overview of the circulating strains. With these advantages, this method is a convenient alternative to the expensive and time-consuming sequence typing method. An important characteristic of the algorithm is the possibility to detect novel types as shown by the characterization of the three isolates with rare prn types.

The algorithm used in this work is based on two amplification reactions using the primer pairs prnC-prnD and prnE-prnD. The former was mainly used for the differentiation of known prn types, with the latter providing additional discrimination and/or confirmation. Due to the presence of point mutations in the prnC primer binding site (Fig. 1A and B), it was not possible to fully optimize the PCR test. The presence of aspecific bands for some rare isolates demonstrates that the PCR conditions are highly critical. Whenever these aspecific bands are observed, performing the described real-time PCR or sequencing of the prn gene is necessary to confirm the observed types.

The primer pairs prnC-prnD and prnE-prnD were specifically designed for the differentiation of prn1 to prn4, the only types that were known at the initiation of the study. Using both
amplifications, we succeeded in differentiating the 11 types known at present into seven groups. These groups were further differentiated by an additional real-time PCR method. Since the difference in the number of repeat units results in a size difference of only 15 bp, we designed a PCR in which short primer pairs prnC-prnD (A) and prnE-prnD (B).

Our study demonstrated that 98% of the Belgian isolates belong to the prn types prn1 to prn3. As much as 9% still belong to the prn1 type, the prn type of the strains that are included in the conventional whole-cell vaccine. Since no isolates are available for the prevaccination period in Belgium, no comparison could be performed for this period.

This method is convenient for large-scale screening of all available isolates in a short time rather than typing a fraction of the collected isolates when sequencing is performed. In the future, application of this method—which is faster and less laborious than sequencing—to other genetic loci such as ptxS1, ptxS3, and tcfA will improve the monitoring of the strains. Finally, this method will also be a useful tool for the collection of typing data from the pre- and post-acellular vaccine eras.

ACKNOWLEDGMENT

This work was supported by OZR grant 624 (Vrije Universiteit Brussel, Brussels, Belgium).

REFERENCES