JCM
Journals.ASM.org

Studying *Bordetella pertussis* Populations by Use of SNPeX, a Simple High-Throughput Single Nucleotide Polymorphism Typing Method

Anne Zeddeman, Sandra Witteveen, Marieke J. Bart, Marjolein van Gent, Han G. J. van der Heide, Kees J. Heuvelman, Leo M. Schouls, Frits R. Mooi

Centre for Infectious Disease Control (CIb), National Institute of Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Large outbreaks of pertussis occur despite vaccination. A first step in the analyses of outbreaks is strain typing. However, the typing of *Bordetella pertussis*, the causative agent of pertussis, is problematic because the available assays are insufficiently discriminatory, not unequivocal, time-consuming, and/or costly. Here, we describe a single nucleotide primer extension assay for the study of *B. pertussis* populations, SNPeX (single nucleotide primer extension), which addresses these problems. The assay is based on the incorporation of fluorescently labeled dideoxynucleotides (ddNTPs) at the 3′ end of allele-specific poly(A)-tailed primers and subsequent analysis with a capillary DNA analyzer. Each single nucleotide polymorphism (SNP) primer has a specific length, and as a result, up to 20 SNPs can be determined in one SNPeX reaction. Importantly, PCR amplification of target DNA is not required. We selected 38 SNPeX targets from the whole-genome sequencing data of 74 *B. pertussis* strains collected from across the world. The SNPeX-based phylogenetic trees preserved the general tree topology of *B. pertussis* populations based on whole-genome sequencing, with a minor loss of details. We envisage a strategy whereby SNP types (SnpTs) are quickly identified with the SNPeX assay during an outbreak, followed by whole-genome sequencing (WGS) of a limited number of isolates representing predominant SnpTs and the incorporation of novel SNPs in the SNPeX assay. The flexibility of the SNPeX assay allows the method to evolve along with the pathogen, making it a promising method for studying outbreaks of *B. pertussis* and other pathogens.

Large outbreaks of pertussis or whooping cough continue to occur, despite vaccination. One of the first steps to resolve the causes of infectious disease outbreaks is strain typing. Strain typing allows for placement of the outbreak in a larger geographical and temporal context, which is required to answer questions related to the origin, evolution, transmission, and characteristics of the outbreak strains. Strain typing can identify successful lineages from which isolates can be selected for more in-depth analysis. For example, whole-genome sequencing (WGS) can be used to reveal genetic loci that may have contributed to the fitness of the successful lineages. Thus, such epidemiologic studies depend on the availability of reproducible typing methods that discriminate between isolates at a desirable level (1, 2). The typing of *Bordetella pertussis*, the causative agent of pertussis, is problematic due to its monomorphic character (3–7). The currently used typing methods for *B. pertussis* include pulsed-field gel electrophoresis (PFGE), multilocus variable-number tandem-repeat analysis (MLVA), multiantigen sequence typing (MAST), and WGS (5, 8–11). WGS is the gold standard but is too expensive for use in routine surveillance. Further, the extraction of relevant data from sequenced genomes requires relatively complex software and expert knowledge. PFGE is currently the most broadly used method to characterize *B. pertussis*. It is highly discriminative but laborious and difficult to standardize (12). Furthermore, it does not always reflect true genetic relationships, because variation may be introduced by recombination between insertion sequence elements, a process that can be reversed with relatively high frequency (6). MLVA, which is PCR based and highly reproducible, is also broadly used, which allows comparisons of strains between different laboratories. However, MLVA is based on variations of only six loci, limiting its discriminatory power (6, 13). For instance, several studies have shown that in the last 20 years, 46 to 74% of the strains belong to two MLVA types (MT), MT-27 and MT-29 (6, 12, 13). Further, two major *B. pertussis* lineages characterized by different alleles for the pertussis toxin promoter, *ptxP1* and *ptxP3*, are not always separated by MLVA (6, 12). Finally, like PFGE, MLVA typing does not always reflect true genetic relationships, as polymorphisms are generated by a reversible variation in the number of tandem repeats (6). MAST targets the alleles of five genes encoding *B. pertussis* proteins used in acellular vaccines (ACVs) (10, 14), which limits its discriminatory power (12). The advantage of MAST, however, is that it gives unequivocal results and allows the identification of vaccine escape strains (7, 14–17). The discriminatory power of MLVA and MAST can be increased by combining the two techniques (11, 12).

The ideal typing method for molecular epidemiologic studies should be unequivocal, simple to use, have a high discriminatory power, and allow high-throughput analysis. For *B. pertussis*, it is also considered valuable to monitor the changes in the alleles of

Received 20 October 2014 **Returned for modification** 22 October 2014
Accepted 28 December 2014

Accepted manuscript posted online 7 January 2015

Citation Zeddeman A, Witteveen S, Bart MJ, van Gent M, van der Heide HGJ, Heuvelman KJ, Schouls LM, Mooi FR. 2015. Studying *Bordetella pertussis* populations by use of SNPeX, a simple high-throughput single nucleotide polymorphism typing method. J Clin Microbiol 53:838–846. doi:10.1128/JCM.02995-14.

Editor: M. J. Loeffelholz

Address correspondence to Frits R. Mooi, frits.mooi@rivm.nl.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.02995-14.

doi:10.1128/JCM.02995-14

The authors have paid a fee to allow immediate free access to this article.

TABLE 1 Dutch bacterial isolates subjected to SNPeX analysis[a]

Yr of isolation by type of isolate	No. with allele profile: *ptxP4 ptxA5 fim3.1*	*ptxP2 ptxA4 fim3.1*	*ptxP6 ptxA2 fim3.1*	*ptxP10 ptxA1 fim3.1*	*ptxP11 ptxA1 fim3.1*	*ptxP13 ptxA1 fim3.1*	*ptxP1 ptxA1 fim3.1*	*ptxP1 ptxA1 fim3.4*	*ptxP1 ptxA2 fim3.1*	*ptxP3 ptxA1 fim3.1*	*ptxP3 ptxA1 fim3.2*	*ptxP3 ptxA1 fim3.3*	Total no.
Diverse (nonrandom)[b]													
1949–1955	1	9					1		8				19
1956–1960									3				3
1961–1965							3		3				6
1966–1970							11		2				13
1971–1975							1						1
1976–1980			1				3		1				5
1981–1985							11		3				14
1986–1990		1		1			15		3	2			22
1991–1995		1					10		2	2			15
1996–2000					1		5		1	2	4		13
2001–2005											1		1
2006–2010						2	3			4	1	1	11
Total	1	11	1	1	1	2	63		26	10	6	1	123
Recent (random)[c]													
2008							1			1			2
2009										6			6
2010										8			8
2011										8			8
2012							1	1		9			11
Total							2	1		32			35

[a] This table lists the Dutch isolates used for SNPeX analysis and shows the isolation period or year, as well as the alleles for the pertussis toxin promoter (*ptxP*), the pertussis toxin A subunit (*ptxA*), pertactin (*prn*), and the serotype 3 fimbrial subunit (*fim3*).

[b] Diverse set of isolates with 83 different MLVA types (not shown) and 11 *ptxP-ptxA-fim3* allele types, which were isolated between 1949 and 2010. The isolates were grouped by periods of 5 years in which the strains were isolated.

[c] In order to investigate the characteristics of isolates that circulated during an epidemic that occurred in 2012, a set of 35 randomly selected recent isolates were subjected to SNPeX analysis. All available strains from 2011 were included, as well as all available strains from January to March 2012.

the components of acellular vaccines. Here, we describe a single nucleotide primer extension assay called SNPeX (single nucleotide primer extension), which complies with these criteria (i.e., is highly discriminatory, gives unequivocal results, is simple, and allows for high-throughput analyses). The assay is based on the incorporation of fluorescently labeled dideoxynucleotides (ddNTPs) into the 3′ end of an allele-specific extension primer with a distinct length. Fragment analysis of the extension products is performed with a capillary DNA analyzer used in many laboratories. In this study, we present a SNPeX target set of 38 predefined single nucleotide polymorphisms (SNPs) that were selected from the WGS data of 74 worldwide strains. Importantly, the SNPeX targets were incorporated into two multiplex SNPeX assays, which were optimized for genomic DNA, without a PCR step. The flexibility of the SNPeX assay allows the method to evolve along with the pathogen, making it a promising method for studying outbreaks of *B. pertussis* and other pathogens.

MATERIALS AND METHODS

SNP selection. Shifts in MLVA types and in the virulence-associated alleles *ptxP*, *ptxA*, *prn*, and *fim3* have been widely used to study the population dynamics of *B. pertussis* (reviewed in references 3 and 18). Therefore, the SNPs for the SNPeX assay were identified by the comparison of 74 genomes, the strains of which represent a diverse collection based on these polymorphic alleles. The strains were isolated between 1949 and 2010 in Africa, Europe, North America, and South America (5). Based on these SNPs, a maximum parsimony tree was constructed using the Mix program from the PHYLIP package (19) (see Fig. S1 in the supplemental material). Supplemented with the tree, the program generated a table that lists the node divisions, along with the accountable SNP or multiple SNPs (see Table S1 in the supplemental material). From this list, the SNP targets for the SNPeX assay were selected. The selection criteria included nonconvergence, the occurrence of an SNP in at least two strains, and a minimum distance of 50 bp between the preceding and following SNP position. In addition, the region containing the SNP position was chosen to be part of the core genome, as described previously (20). If multiple SNPs were available for a particular node division, a position was chosen at random. Further, at least one SNP was selected that is specific for the *ptxP*, *ptxA*, *ptxC*, and *fim3* alleles, because these alleles have been included in many previous studies (see Table S2 in the supplemental material).

***B. pertussis* isolates and culture conditions.** The bacteria were cultured for 3 to 4 days at 35°C on Bordet-Gengou (BG) agar plates that were supplemented with 15% sheep blood. *B. pertussis* strain B1917 was used to test the stability of the SNPeX targets by subculturing it 53 times. The colonies that were transferred 1, 25, and 53 times were typed with SNPeX. Two strains, *B. pertussis* B1203 and B1690, were analyzed twice from independently grown colonies in order to test the reproducibility of the SNPeX assay.

Two sets of isolates were analyzed using the SNPeX assay. The first set, collected from 1949 to 2010, contained 123 Dutch isolates that represented 83 MLVA types and 11 *ptxP-ptxA-fim3* allele types (Table 1). In addition, 35 randomly selected Dutch isolates from 2008 to 2012 were added to this set. These included 11 isolates that were isolated early (Jan-

uary to March) during a Dutch pertussis epidemic in 2012 (see Table S3 in the supplemental material). The second and more geographically diverse set comprised 60 strains from six continents and 16 countries (see Table S3).

DNA isolation. Genomic DNA was isolated from *B. pertussis* isolates using the GenElute bacterial genomic DNA kit (Sigma-Aldrich, Zwijndrecht, the Netherlands), according to the manufacturer's instructions for Gram-negative bacteria. The minimum concentration of DNA required for the SNPeX assay was 200 ng/μl. If possible, in order to achieve a minimum DNA concentration of 200 ng/μl, the extraction buffer was evaporated by heating open reaction tubes at 40°C for 30 min to 2 h.

Multiantigen sequence typing. For MAST, the alleles for the pertussis toxin promoter (*ptxP*), regions 1 and 2 of the pertactin gene (*prn*), and the genes for serotype 2 fimbriae (*fim2*) and serotype 3 fimbriae (*fim3*) were sequenced as described previously (10, 21).

Multilocus variable-number tandem-repeat analysis. The number of tandem repeats in six loci (VNTR1, VNTR3A, VNTR3B, VNTR4, VNTR5, and VNTR6) was determined as described before (11).

SNPeX preview assay. Allele-specific primers were designed on the *B. pertussis* Tohama I genome sequence (National Center for Biotechnology Information [NCBI] reference sequence accession no. NC_002929) using the CLC Main Workbench 7.0.3 software (CLC bio, Aarhus, Denmark). The primers were purified by polyacrylamide gel electrophoresis (Eurogentec, Maastricht, the Netherlands). To assess the mobility of the four possible fluorescently labeled extension products of each primer, the products were visualized with a single-plex SNPeX preview assay using the SNaPshot primer focus kit (Life Technologies Corporation, CA, USA). The SNPeX preview assay was used to adjust the mix of primers to a suitable combination to prevent overlapping peaks (see Results). A SNPeX preview reaction mixture contained 0.1 μl of primer focus enzyme, 4 μl of primer focus ddNTP mix, 1 μl of 10× primer focus reaction buffer, 1 μl of SNaPshot primer focus cofactor (Life Technologies Corporation), 1 μl of primer (2 pmol/μl), and 2.9 μl of water. The reaction mixture was heated to 37°C for 15 min, followed by 70°C for 10 min. In order to remove unincorporated ddNTPs, 1 μl of shrimp alkaline phosphatase (SAP) was added to the reaction mixture and incubated for 60 min at 37°C and 15 min at 72°C. Subsequently, 2 μl of reaction product was dissolved in 10 μl of GeneScan 120 LIZ size standard (100× diluted) (Life Technologies Corporation) and denatured at 95°C for 5 min. After denaturation, the samples were run on a 48-capillary ABI Prism 3730 DNA analyzer (Applied Biosystems, CA, USA), using the POP-7 polymer and GeneMapper50_POP7_1 module.

SNPeX assay. The primers were dissolved into two reaction mixtures to a concentration of 0.5 to 4 pmol/μl per primer. The concentration of the primer was adjusted to reach a fluorescence signal of 2,000 to 4,000 relative fluorescence units (RFUs) in the assay (Table 2). The multiplex SNPeX assay was adjusted from that of Vallone et al. (22), using the ABI Prism SNaPshot multiplex kit (Life Technologies Corporation, CA, USA) according to the manufacturer's instructions, with slight modifications. Briefly, 4 μl of genomic DNA (>200 ng/μl) was added to 1 μl of primer mix. This mixture was denatured in a preheated thermal cycler at 95°C for 2 min, after which it was promptly put on ice and spun down. Five microliters of cold SNaPshot reaction mix (Life Technologies Corporation) was added to the sample, and the primer extension reaction was started, with denaturation at 95°C for 2 min, followed by 100 cycles of 96°C for 10 s and 61°C for 30 s. Unincorporated fluorescently labeled ddNTPs were removed by adding 1.5 μl (1 U/μl) of SAP to the mixture, followed by incubation for 60 min at 37°C. Afterwards, the enzyme was inactivated by incubation for 15 min at 80°C. Subsequently, 2 μl of purified SNaPshot reaction product was added to 10 μl of LIZ 120 size standard (100× diluted) (Applied Biosystems, CA, USA). The samples were denatured for 5 min at 95°C, cooled down to room temperature, and run on a 48-capillary ABI Prism 3730 DNA analyzer (Applied Biosystems). The GeneMapper50_POP7_1 run module was adjusted to a prerun time of 120 s, an injection volume of 5.0 kV, injection time of 20 s, data delay of 400 s, and run time of 650 s. Nucleotide incorporation was analyzed using GeneMarker 1.5.1 (Softgenetics, PA, USA).

Data analysis. Peak calling was performed using the software GeneMarker 1.5.1 (Softgenetics, State College, PA, USA), and the profiles were imported into BioNumerics 7.1 (Applied Maths, Sint-Martens-Latem, Belgium) as the character data and minimum spanning trees (MSTs) with hypothetical nodes were created. Each SNP position was treated as an independent categorical character. The tree was rooted with strains carrying the *ptxP4* allele, as previous work showed that *ptxP4* strains are most closely related to *Bordetella bronchiseptica*, a species from which *B. pertussis* has evolved (5, 23). Ridom EpiCompare version 1.0.6 was used to calculate the discriminatory index (DI) and confidence interval (CI) of the SNPeX assay (Ridom Bioinformatics, Münster, Germany).

RESULTS

Selection of SNPeX targets from whole-genome sequences. A comparison of 74 worldwide genomes revealed a total of 2,811 SNPs that were present in all strains (5) (see Table S4 in the supplemental material). An analysis of the maximum parsimony tree constructed with these SNPs, along with our selection criteria, resulted in 38 SNP targets (Table 2).

***In silico* evaluation of SNPeX targets.** The 38 SNPeX targets were extracted *in silico* from the 74 genomes described above. The tree based on 2,811 SNPs derived from whole-genome sequences (Fig. 1A) was compared to the tree based on the selected 38 SNPeX targets (Fig. 1B; see also Fig. S2 in the supplemental material). With two exceptions (strains B3197 and B0585), the tree based on 2,811 SNPs distinguished all strains, resulting in 73 SNP types. The clustering of the same strains based on the 38 SNPs targeted by the SNPeX resulted in 34 SNPeX types (SnpTs). As expected, a reduction in the amount of SNPs used for clustering resulted in the coalescence of the branches. Importantly, the general topology (main branches) of the tree based on 2,811 SNPs was conserved. This was demonstrated by the conservation of the relative positions of most SnpTs in the SNPeX tree compared with the tree based on 2,811 SNPs (Fig. 1A). However, not all relationships between the strains were preserved. For example, in the tree that resulted from the 2,811 SNPs, SnpT30 was centered between SnpT26, SnpT28, SnpT29, SnpT32, and SnpT33, but in the tree based on the selected 38 SNPeX targets, SnpT30 was directly connected to SnpT31 and SnpT29 only. The discriminatory indices of the complete whole-genome sequence SNP set and the SNPeX set were 1.00 (95% CI, 0.999 to 1.0) and 0.974 (95% CI, 0.965 to 0.983), respectively. Therefore, further SnpT designations were based on this SNPeX set.

Setup and principle of the method. Figure S3 in the supplemental material shows the principle of the SNPeX assay. Primers were designed that bind adjacent to each SNP position (see Table S5 in the supplemental material). Each primer had a distinctive length (35 to 100 bp) due to a variable poly(A) tail. SNP typing was performed directly on DNA without an amplification step. For SNP identification, the primer was elongated at the targeted SNP with a fluorescently marked ddNTP, using fluorescent dyes specific for the four nucleotides. With this approach, the length of the elongation product defined the SNP position, whereas the color of the fluorescent nucleotide indicated the base at the SNP position. The sizing and determination of the color of a fragment is referred to as peak calling. Further, the variable length due to the poly(A) tail allowed for multiplexing.

Before the SNPeX assay could be applied to *B. pertussis* isolates, the suitability of each of the selected primers for multiplexing was

TABLE 2 SNPeX targets

SNP no.[a]	Position[b]	Locus ID for SNP	Gene[c]	aa change caused by nonsilent SNPs	SNP specific for[d]:	nt of Tohama at SNP	Variant at SNP position	Primer length (bases)[e]	Primer concn in multiplex mix (pmol/μl)
1	T2905672	BP2738	*bapC*	Silent	*ptxP4*	A	G	92	1
2	T2678045	BP2526	—	S410L	*ptxP6*	G	A	96	1
3	T3650318	BP3439	*dnt*	G1079S	*ptxP2*	G	A	88	4
4	T0824268	Intergenic			*ptxP1*	C	T	77	2
5	T1774569	BP1690	—	Silent	*ptxP1*	T	C	82	2
6	T2244171	BP2120	*gor*	A341T		G	A	96	1
7	T2166028	Intergenic			*ptxP1*	C	T	53	2
8	T2277180	BP2156	—	Silent	*ptxP1*	G	A	94	2
9	T0835998	Intergenic				G	A	64	2
10	T0063037	BP0063	—	Silent	*ptxP1*	C	T	46	2
11	T3988941	BP3783	*ptxA*	M228I		G	A	60	2
12	T1931433	Intergenic				A	G	36	2
13	T3686708	BP3474	—	S450L	*ptxP1*	G	A	66	2
14	T0220937	BP0215	*ppc*	Silent		G	A	72	4
15	T3764370	BP3553	*risS*	Silent	*ptxP1*	G	A	94	2
16	T0514171	Intergenic				G	A	86	2
17	T0511992	Intergenic				A	G	74	2
18	T1362787	BP1291	—	S37N		G	A	50	4
19	T2587815	Intergenic				T	C	90	0.5
20	T1107425	BP1062	—	M258I		C	T	82	2
21	T2615461	Intergenic				G	A	100	2
22	T3240976	BP3040	*bllY*	Silent	*ptxP1*	G	A	84	2
23	T0072405	BP0074	*htpG*	R257S	*ptxP1*	C	A	90	1
24	T1702741	BP1619	—	Silent		C	T	74	4
25	T3085804	Intergenic				G	T	84	0.5
26	T3568653	BP3346	—	R198C		G	A	52	2
27	T3988168	BP3783P	*ptxP*			G	A	98	2
28	T3991376	BP3787	*ptxC*	Silent	*ptxP3 ptxP15 ptxP18*	C	T	68	2
29	T3161156	BP2974	—	L118F	*ptxP3*	G	A	88	0.5
30	T1547488	BP1471	—	N2S	*ptxP3 ptxP15 ptxP18*	A	G	80	1
31	T3833330	BP3618	*rplV*	Silent	*ptxP3*	C	A	50	2
32	T2374322	BP2249	*bscI*	Y114C	*ptxP3 ptxP15 ptxP18*	T	C	35	2
33	T0667028	BP0658	—	Q30	*ptxP3*	G	A	68	2
34	T0838886	BP0814	—	Silent	*ptxP3*	G	A	42	2
35	T2681216	BP2528A	—	Silent	*ptxP3*	G	A	72	4
36	T1647861	BP1568	*fim3*	A87E	*ptxP3 ptxP15 ptxP18*	C	A	35	2
37	T2505238	BP2366	*prpB*	V208A	*ptxP3 ptxP15 ptxP18*	T	C	60	1
38	T0860069	Intergenic			*ptxP15 ptxP18*	C	T	80	2

[a] SNP numbers are as presented in Fig. 1B.
[b] Position of the targeted SNP in strain Tohama I. The numbering of the SNP site position is based on the plus strand of the genome (accession no. NC_002929) and starts from the origin of replication. The extension primer sequences can be found in Table S5 in the supplemental material.
[c] Name of the gene in which the SNP is located. —, location unknown.
[d] *ptxP* allele(s) uniquely associated with this SNP.
[e] Length of the primer [specific part plus poly(A) tail] that targets the SNP.

assessed using a SNPeX preview reaction. Figure S4 in the supplemental material shows the principle of the SNPeX preview assay, using images of three primers used in this study. The primer focus enzyme in the SNPeX preview assay extends the primer with all four labeled nucleotides in a single-plex reaction. Note that there is no *B. pertussis* (or any other) DNA template present in this reaction. Subsequently, all potential extension products can be identified by capillary electrophoresis, and therefore, it can be

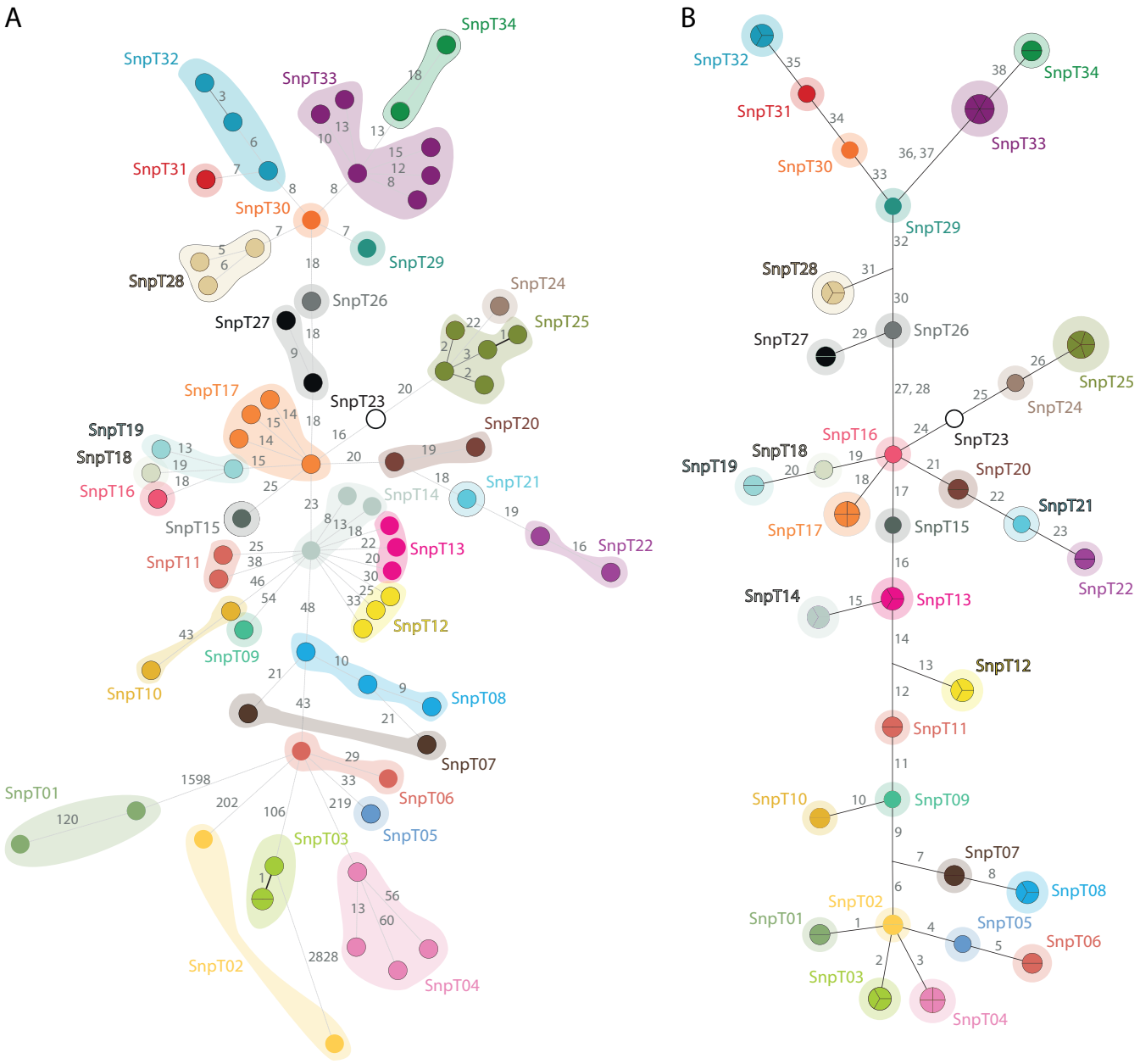

FIG 1 Minimum spanning tree of 74 strains based on 2,811 SNPs (A) and the *in silico*-derived 38 SNPeX targets (B). (A) Colored circles and shading define sequence types (SnpTs), as identified by the SNPeX assay. The numbers in the connecting lines refer to the number of SNPs distinguishing the connected strains, and the lengths of the lines are logarithmically scaled to the number of SNPs. (B) Segments of circles indicate the number of strains that merged. The numbers on the connecting lines refer to the SNP number distinguishing the connected strains (as in Table 2), and the length of the lines is scaled to the number of SNPs. The whole-genome sequences of these were published previously (5). SnpT, SNPeX sequence type.

verified that the products of different primers do not overlap in mobility. We designed the multiplex to have 19 and 20 primers in each mix. One primer appeared problematic in the SNP preview analysis because of overlapping peaks with other primers. We removed it from the first multiplex mix. The two mixes were incorporated into the SNPeX assay, which was performed directly on DNA without a PCR amplification step. The reactions were carried out in microtiter plates, allowing the analysis of 47 samples per plate (always including one reaction without template DNA as negative control). The costs of the SNPeX assay were approximately $6.00 per sample. An example of the multiplex SNPeX results of three strains and three primers is presented in Fig. S5 in the supplemental material.

Turnaround time and success rate. Turnaround time is an important feature for a typing assay, especially when it concerns urgent pathogen typing during an epidemic. In this experimental design, a single person should be able to prepare 30 *B. pertussis* DNA samples per day from cells grown on agar plates. To carry out the SNPeX reaction, purify the fragments, and run the fragments takes 70, 75, and 30 min, respectively, per 96-well plate. Thus, after DNA isolation, it is possible to prepare and run two 96-well plates (equivalent to 95 strains and 38 SNPs) within 5 h.

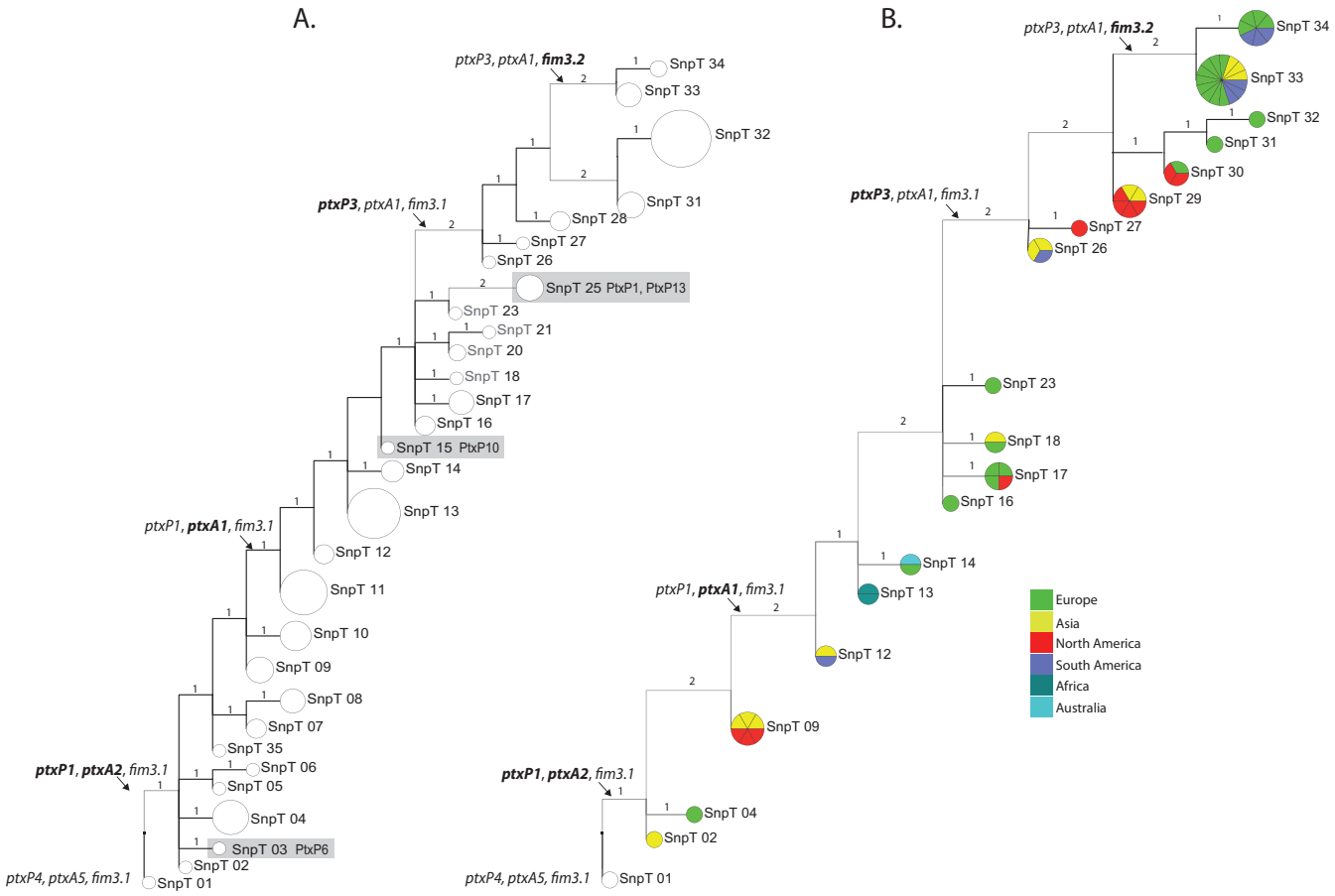

FIG 2 Rooted minimum spanning trees of Dutch (A) and global (B) *B. pertussis* isolates based on the SNPeX assay and 38 SNPs. The *ptxP*, *ptxA*, and *fim3* alleles are shown next to the branches. The alleles in bold type indicate their earliest appearance in the tree. The numbers refer to the distance (in SNPs) between the branches, and the circle sizes are scaled to the number of isolates in each SnpT. (A) One hundred fifty-eight Dutch strains were divided into 30 SnpTs. The tree separately clusters the *ptxP1* and *ptxP3* isolates. Further, within the *ptxP3* clade, *fim3.1* and *fim3.2* isolates were found in distinct branches. This is a well-established topology for *B. pertussis* strains based on WGS. SnpTs associated with *ptxP6*, *ptxP10*, and *ptxP13* alleles are shaded. (B) Sixty geographically diverse strains were divided into 19 SnpTs. The strains do not cluster geographically (the continents are indicated by color). The tree is also similar to trees based on WGS (4, 5, 24).

The *in silico* analyses required 3 h per run and included verification of the assignment of a nucleotide to an allele (peak calling, see Fig. S4 in the supplemental material) and preparation of an output file that lists the SNPs and allows secondary analyses (e.g., clustering). We were able to process four multiplex plates, equivalent to 192 strains, per working day (not including DNA isolation).

In total, the SNPeX reaction was performed on 260 samples, of which 200 samples directly revealed a complete SNP set (76% success rate). Most samples with an incomplete SNP set had a DNA concentration of <200 ng/μl. One extra run was performed on the samples that had an incomplete set of SNPs, which resulted in 226 samples with a complete SNP set (86% success rate). A low DNA concentration was the main reason for the incomplete SNP sets.

Reproducibility and reliability. Throughout the 1×, 25×, and 53× subculturing of a strain (B1917), all SNPs were identical (results not shown). Also, two strains were analyzed twice from independently grown colonies. The SNPs of both samples from a single isolate were identical (results not shown). In addition, 23 of the strains that were subjected to this SNPeX assay were also analyzed with WGS from independent cultures. The SNPs identified by the SNPeX assay were checked with the whole-genome data, and no discrepancies were found (results not shown).

Application of the SNPeX assay to type locally and globally collected isolates. SNPeX typing of 158 Dutch strains (isolated from 1949 to 2012) revealed 30 SnpTs (Fig. 2A; see also Table S3 in the supplemental material), with a diversity index of 0.912 (95% CI, 0.89 to 0.934). The isolates from the 2012 Dutch epidemic ($n = 11$) belonged to SnpT12 ($n = 1$), SnpT13 ($n = 1$), SnpT31 ($n = 1$), and SnpT32 ($n = 8$). SnpT32 predominated during the 2012 epidemic but also in previous years since 2009 (Fig. 3), which suggests that either the 2012 epidemic was not associated with the expansion of novel strains, or the SNPeX assay did not contain the SNPs that are characteristic of the 2012 strains. Preliminary WGS data showed that the strains from 2012 were characterized by two SNPs (reference positions in the Tohama I genome, 2570058 and 2570056), both located in a pseudogene (BP2427) (results not shown). The usefulness of the SNPeX assay was also explored using a more geographically diverse collection of *B. pertussis* strains (Fig. 2B). SNPeX typing clustered the strains into 19 SnpTs, with a diversity index of 0.893 (95% CI, 0.85 to 0.94). As observed previously (5), there was no evidence of geographic clustering.

The trees of the local (Dutch) and global strains (Fig. 2A and B, respectively) had similar features, such as distinct branches for

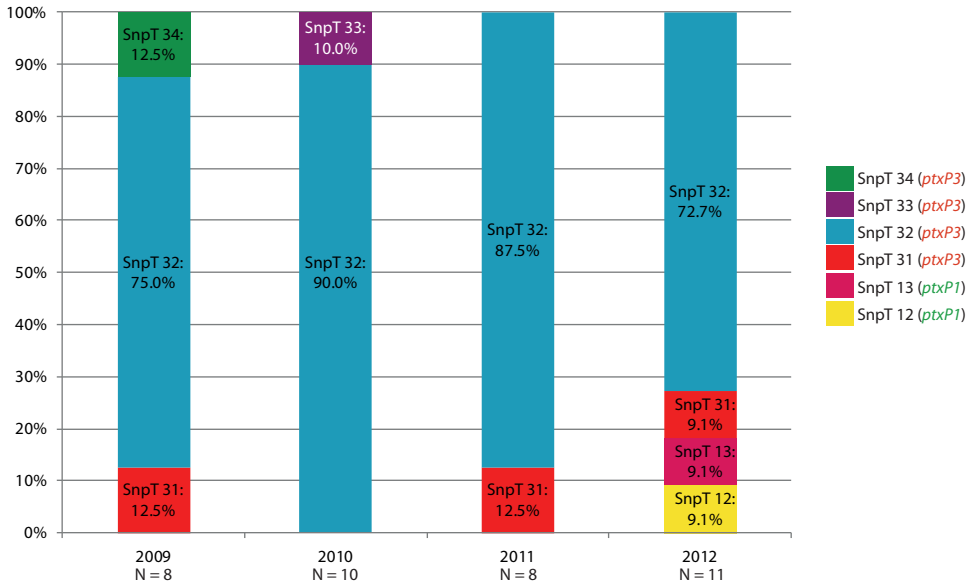

FIG 3 Distribution of SnpT frequency of the Dutch strains from 2009 to 2012. The frequency of the SnpTs in each year is shown; next to the isolation year of the isolates on the *x* axis, the number of isolates is given for each isolation year. SnpT31 was most prevalent, and no shift in SnpT prevalence was observed preceding or during the Dutch pertussis epidemic in 2012. All 37 Dutch strains analyzed with the SNPeX assay and isolated from 2009 to 2012 are included in this analysis.

isolates carrying the *ptxP1/ptxP3* and *fim3.1/fim3.2* alleles. Further, consistent with previous work (4, 5, 16, 24), both trees revealed that the mutations in the *ptxA*, *ptxP*, and *fim3* genes accumulated clonally. Progressing up the tree from the root, the *ptxP* allele generally changed from *ptxP4* to *ptxP1* to *ptxP3*, and the *ptxA* allele changed from *ptxA5* to *ptxA2* to *ptxA1*. In both trees, the *ptxP3* clade was divided into two main branches that coincided with the *fim3.1* and *fim3.2* alleles. Branches characterized by distinct *ptxP*, *ptxA*, and *fim3* alleles in the phylogenetic trees based on 2,811 SNPs from WGS analysis were retained in the tree based on the 38 SNPeX targets (see Fig. S2 in the supplemental material). To summarize, the trees based on only the 38 SNPs were very similar to previously published trees based on WGS (4, 5, 24).

Combining the SNPeX assay with MLVA. MLVA has been widely used to compare *B. pertussis* strains, although it does not accurately describe genetic relationships due to the reversibility of the number of repeats (6). However, MLVA may be useful for subdividing the SnpTs. Therefore, the MLVA type of each Dutch strain carrying the *ptxP3* allele ($n = 49$) was determined, followed by hierarchical clustering (Fig. 4). The strains were first clustered using the SNPeX assay, and subsequently, the SnpTs were subdivided based on MLVA type. This approach resulted in 17 combined SNPeX-MLVA types, of which the discriminatory index was 0.696 (95% CI, 0.55 to 0.84) (compared to a discriminatory indices of 0.58 [95% CI, 0.429 to 0.731] and 0.644 [95% CI, 0.49 to 0.797] based on SNPeX and MLVA only, respectively). The combined approach and the resulting tree revealed the convergence of MLVA types. For instance, MT27 was not only associated with the most prevalent SnpT32 but also with SnpT31 and SnpT33. This hierarchical approach establishes relationships between the strains by combining the relatively slow molecular clock based on SNPs together with the faster clock based on the insertion and deletion of repeats (MLVA) (25). Although the combined approach does further subdivide the SnpTs, SnpT32-MT27 was predominant ($n = 27$ [55%]). Also, 8 of 9 isolates from the 2012 epidemic were SnpT32-MT27, so MLVA did not further resolve the typing of these strains.

DISCUSSION

B. pertussis is a highly monomorphic pathogen, and current methods that are used to study *B. pertussis* populations either have low discriminatory power, are difficult to standardize, or do not reveal true genetic relationships due to convergence (6). The ultimate approach to determine the relationships between strains is WGS. However, WGS is still relatively expensive, and the data analysis requires complex software. Here, we present the SNPeX assay, a fast, low-cost, and high-throughput method that types strains based on SNPs. Importantly, the SNPeX assay does not require the amplification of target DNA, greatly simplifying the method. Starting from purified DNA, it was possible to characterize 38 SNPs of 95 strains in 1 day. The assay requires the use of capillary electrophoresis platforms that are common in laboratories or can easily be employed by outsourcing. The method is based on the separation of primer extension products of distinct lengths, allowing for multiplexing up to 20 primers per microtiter plate well (see Fig. S3 in the supplemental material). In a previous study, we selected 87 SNPs for strain typing based on the Tohama I strain and six Dutch clinical isolates (6). The selection of the 38 SNPeX positions targeted in this assay was based on the comparison of the genome sequences of 74 worldwide strains that yielded 2,811 SNPs; therefore, it is more useful for global studies. The two SNP sets overlap by 6 SNPs. The *in silico* comparison of the 38 SNP positions with the 2,811 SNPs derived from WGS revealed trees of very similar topology, despite the >70-fold reduction in the number of SNP positions addressed (Fig. 1; see also Fig. S2 in the supplemental material). For example, both trees showed that the mutations in *ptxA*, *ptxP*, and *fim3* genes accumulated clonally, and important branches were retained in the tree based on the smaller SNP set. The SNPeX assay was applied to a collection of temporally diverse Dutch isolates (isolated from 1949 to 2012) and a

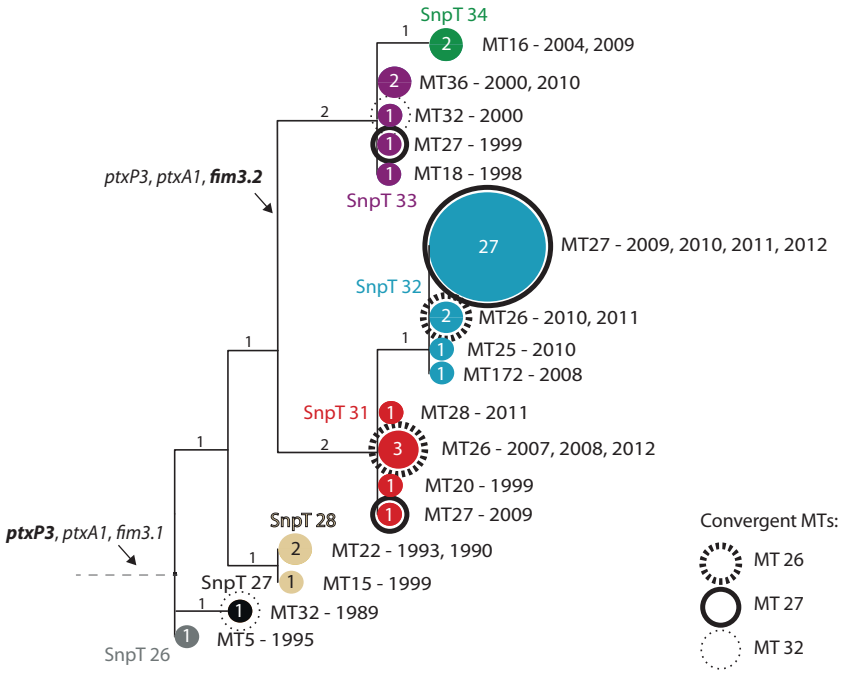

FIG 4 Hierarchical minimum spanning trees of 49 Dutch *ptxP3* strains based on combined SNPeX typing and MLVA. The strains were first clustered based on SNP typing, and subsequently, MLVA types were assigned to the terminal branches. The *ptxP*, *ptxA*, and *fim3* alleles are shown next to the branches. The alleles in bold type indicate their earliest appearance in the tree. The numbers on the connecting lines refer to the distance between the nodes. The circle size is relative to the number of strains with that particular combined SnpT/MT. Most *ptxP3* strains ($n = 27$ [55%]) belong to SnpT32-MT27. Note the occurrence of identical MTs in distinct SnpT branches (convergence, indicated by the extra ring around the SnpTs). Convergence suggests highly reversible mutations or strong selection. In both cases, such variable loci are less suitable for establishing relationships between strains. MT, MLVA type.

collection of geographically diverse worldwide isolates (from six continents and 16 countries, isolated between 1986 and 2010). In both cases, the trees showed the clonal accumulating changes in the *ptxA*, *ptxP*, and *fim3* alleles. These allele patterns were also observed in the tree based on 2,811 SNPs from WGS (see Fig. S2) and have been considered to contribute to the persistence of pertussis (reviewed in references 3 and 18; see also reference 24). A limitation of this assay is that novel SNPs are not detected. SNPeX analyses of the strains isolated during the 2012 Dutch epidemic, in which SNPeX analysis did not reveal novel SnpTs compared to those of previous years, exemplifies this. The WGS analyses of these strains identified two unique SNPs associated with the 2012 strains, both which were present in a pseudogene (BP2427). In order to distinguish the 2012 strains, one of these could be selected and added to the SNPeX assay. Within the current primer length range, at least two primers could theoretically be added to each mix (e.g., with a total primer length of 38 or 56 bases in mix one, or total primer length of 40, 44, 64, or 76 bases in mix two). However, this scheme should be tested empirically. This exemplifies the SNPeX the strategy we envisage, i.e., the identification of predominant SnpTs during an outbreak, followed by WGS of a limited number of representatives of this SnpT. Subsequently, the SNPeX targets can be updated by WGS. SNPs may be added to, or removed from, the SNPeX assay as *B. pertussis* evolves. We have used a primer length range of 35 to 100 bases, but with the LIZ 120 marker that was used, it is theoretically possible to extend to a primer length of 120 bases, allowing the assay to incorporate another 20 primers (as differences in primer length should be at least two bases). Importantly, the data from the SNPeX assay can be compared to or complemented with the approximately 350 *B. pertussis* whole-genome sequences available by *in silico* extraction of SNPs. Here, the assay was applied directly to DNA isolated from cultures. However, the inclusion of a PCR step that amplifies the region around each SNPeX target would allow the assay to be applied directly to clinical samples. This would require the development of multiplex PCR to limit the workload. The broad implementation of the SNPeX assay will reveal the global spread of successful lineages and enhance our understanding of how *B. pertussis* is able to cause large outbreaks even in highly vaccinated populations.

ACKNOWLEDGMENTS

This project was supported by the National Institute for Public Health and the Environment (RIVM), the Netherlands (SOR project S/230456).

REFERENCES

1. **Struelens MJ.** 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. Clin Microbiol Infect **2**:2–11. http://dx.doi.org/10.1111/j.1469-0691.1996.tb00193.x.
2. **Foxman B, Zhang L, Koopman JS, Manning SD, Marrs CF.** 2005. Choosing an appropriate bacterial typing technique for epidemiologic studies. Epidemiol Perspect Innov **2**:10. http://dx.doi.org/10.1186/1742-5573-2-10.
3. **Mooi FR.** 2010. *Bordetella pertussis* and vaccination: the persistence of a genetically monomorphic pathogen. Infect Genet Evol **10**:36–49. http://dx.doi.org/10.1016/j.meegid.2009.10.007.
4. **Octavia S, Maharjan RP, Sintchenko V, Stevenson G, Reeves PR, Gilbert GL, Lan R.** 2011. Insight into evolution of *Bordetella pertussis* from comparative genomic analysis: evidence of vaccine-driven selection. Mol Biol Evol **28**:707–715. http://dx.doi.org/10.1093/molbev/msq245.
5. **Bart MJ, Harris SR, Advani A, Arakawa Y, Bottero D, Bouchez V, Cassiday PK, Chiang CS, Dalby T, Fry NK, Gaillard ME, van Gent M, Guiso N, Hallander HO, Harvill ET, He Q, van der Heide HG, Heu-**

velman K, Hozbor DF, Kamachi K, Karataev GI, Lan R, Lutylska A, Maharjan RP, Mertsola J, Miyamura T, Octavia S, Preston A, Quail MA, Sintchenko V, Stefanelli P, Tondella ML, Tsang RS, Xu Y, Yao SM, Zhang S, Parkhill J, Mooi FR. 2014. Global population structure and evolution of *Bordetella pertussis* and their relationship with vaccination. mBio **5:**e01074. http://dx.doi.org/10.1128/mBio.01074-14.

6. **van Gent M, Bart MJ, van der Heide HG, Heuvelman KJ, Kallonen T, He Q, Mertsola J, Advani A, Hallander HO, Janssens K, Hermans PW, Mooi FR.** 2011. SNP-based typing: a useful tool to study *Bordetella pertussis* populations. PLoS One **6:**e20340. http://dx.doi.org/10.1371/journal.pone.0020340.
7. **Van Loo IH, Mooi FR.** 2002. Changes in the Dutch *Bordetella pertussis* population in the first 20 years after the introduction of whole-cell vaccines. Microbiology **148:**2011–2018.
8. **Advani A, Hallander HO, Dalby T, Krogfelt KA, Guiso N, Njamkepo E, von Konnig CH, Riffelmann M, Mooi FR, Sandven P, Lutynska A, Fry NK, Mertsola J, He Q.** 2013. Pulsed-field gel electrophoresis analysis of *Bordetella pertussis* isolates circulating in Europe from 1998 to 2009. J Clin Microbiol **51:**422–428. http://dx.doi.org/10.1128/JCM.02036-12.
9. **Mooi FR, Hallander H, Wirsing von Konig CH, Hoet B, Guiso N.** 2000. Epidemiological typing of *Bordetella pertussis* isolates: recommendations for a standard methodology. Eur J Clin Microbiol Infect Dis **19:**174–181. http://dx.doi.org/10.1007/s100960050455.
10. **van Loo IH, Heuvelman KJ, King AJ, Mooi FR.** 2002. Multilocus sequence typing of *Bordetella pertussis* based on surface protein genes. J Clin Microbiol **40:**1994–2001. http://dx.doi.org/10.1128/JCM.40.6.1994-2001.2002.
11. **Schouls LM, van der Heide HG, Vauterin L, Vauterin P, Mooi FR.** 2004. Multiple-locus variable-number tandem repeat analysis of Dutch *Bordetella pertussis* strains reveals rapid genetic changes with clonal expansion during the late 1990s. J Bacteriol **186:**5496–5505. http://dx.doi.org/10.1128/JB.186.16.5496-5505.2004.
12. **Advani A, Van der Heide HG, Hallander HO, Mooi FR.** 2009. Analysis of Swedish *Bordetella pertussis* isolates with three typing methods: characterization of an epidemic lineage. J Microbiol Methods **78:**297–301. http://dx.doi.org/10.1016/j.mimet.2009.06.019.
13. **Litt DJ, Jauneikaite E, Tchipeva D, Harrison TG, Fry NK.** 2012. Direct molecular typing of *Bordetella pertussis* from clinical specimens submitted for diagnostic quantitative (real-time) PCR. J Med Microbiol **61:**1662–1668. http://dx.doi.org/10.1099/jmm.0.049585-0.
14. **Packard ER, Parton R, Coote JG, Fry NK.** 2004. Sequence variation and conservation in virulence-related genes of *Bordetella pertussis* isolates from the UK. J Med Microbiol **53:**355–365. http://dx.doi.org/10.1099/jmm.0.05515-0.
15. **Mooi FR, van Oirschot H, Heuvelman K, van der Heide HG, Gaastra W, Willems RJ.** 1998. Polymorphism in the *Bordetella pertussis* virulence factors P.69/pertactin and pertussis toxin in the Netherlands: temporal trends and evidence for vaccine-driven evolution. Infect Immun **66:**670–675.
16. **van Gent M, Bart MJ, van der Heide HG, Heuvelman KJ, Mooi FR.** 2012. Small mutations in *Bordetella pertussis* are associated with selective sweeps. PLoS One **7:**e46407. http://dx.doi.org/10.1371/journal.pone.0046407.
17. **Octavia S, Sintchenko V, Gilbert GL, Lawrence A, Keil AD, Hogg G, Lan R.** 2012. Newly emerging clones of *Bordetella pertussis* carrying *prn2* and *ptxP3* alleles implicated in Australian pertussis epidemic in 2008–2010. J Infect Dis **205:**1220–1224. http://dx.doi.org/10.1093/infdis/jis178.
18. **Mooi FR, Van Der Maas NA, De Melker HE.** 2014. Pertussis resurgence: waning immunity and pathogen adaptation—two sides of the same coin. Epidemiol Infect **142:**685–694. http://dx.doi.org/10.1017/S0950268813000071.
19. **Felsenstein J.** 2005. PHYLIP version 3.6. Department of Genome Sciences, University of Washington, Seattle, WA.
20. **King AJ, van Gorkom T, Pennings JL, van der Heide HG, He Q, Diavatopoulos D, Heuvelman K, van Gent M, van Leeuwen K, Mooi FR.** 2008. Comparative genomic profiling of Dutch clinical *Bordetella pertussis* isolates using DNA microarrays: identification of genes absent from epidemic strains. BMC Genomics **9:**311. http://dx.doi.org/10.1186/1471-2164-9-311.
21. **Mooi FR, Van Loo IHM, Van Gent M, He Q, Bart MJ, Heuvelman KJ, de Greeff SC, Diavatopoulos D, Teunis P, Nagelkerke N, Mertsola J.** 2009. *Bordetella pertussis* strains with increased toxin production associated with pertussis resurgence. Emerg Infect Dis **15:**1206–1213. http://dx.doi.org/10.3201/eid1508.081511.
22. **Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ.** 2004. A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. Int J Legal Med **118:**147–157. http://dx.doi.org/10.1007/s00414-004-0428-5.
23. **Diavatopoulos DA, Cummings CA, Schouls LM, Brinig MM, Relman DA, Mooi FR.** 2005. *Bordetella pertussis*, the causative agent of whooping cough, evolved from a distinct, human-associated lineage of *B. bronchiseptica*. PLoS Pathog **1:**e45. http://dx.doi.org/10.1371/journal.ppat.0010045.
24. **Lam C, Octavia S, Bahrame Z, Sintchenko V, Gilbert GL, Lan R.** 2012. Selection and emergence of pertussis toxin promoter *ptxP3* allele in the evolution of *Bordetella pertussis*. Infect Genet Evol **12:**492–495. http://dx.doi.org/10.1016/j.meegid.2012.01.001.
25. **Dimovski K, Cao H, Wijburg OL, Strugnell RA, Mantena RK, Whipp M, Hogg G, Holt KE.** 2014. Analysis of *Salmonella enterica* serovar Typhimurium variable-number tandem-repeat data for public health investigation based on measured mutation rates and whole-genome sequence comparisons. J Bacteriol **196:**3036–3044. http://dx.doi.org/10.1128/JB.01820-14.