PLoS one

SNP-Based Typing: A Useful Tool to Study *Bordetella pertussis* Populations

Marjolein van Gent[1][9], **Marieke J. Bart**[1][9], **Han G. J. van der Heide**[1], **Kees J. Heuvelman**[1], **Teemu Kallonen**[2,3], **Qiushui He**[2], **Jussi Mertsola**[2], **Abdolreza Advani**[4], **Hans O. Hallander**[4], **Koen Janssens**[5], **Peter W. Hermans**[6], **Frits R. Mooi**[1]*

1 Laboratory for Infectious Diseases and Screening, Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands, 2 Department of Infectious Disease Surveillance and Control, National Institute for Health and Welfare, Turku, Finland, 3 Turku Graduate School for Biomedical Sciences, University of Turku, Turku, Finland, 4 Department of Bacteriology, Swedish Institute for Infectious Disease Control, Solna, Sweden, 5 Applied Maths NV, Sint-Martens-Latem, Belgium, 6 Laboratory of Pediatric Infectious Diseases, Radboud University Nijmegen Medical Center, Radboud University Nijmegen, Nijmegen, The Netherlands

Abstract

To monitor changes in *Bordetella pertussis* populations, mainly two typing methods are used; Pulsed-Field Gel Electrophoresis (PFGE) and Multiple-Locus Variable-Number Tandem Repeat Analysis (MLVA). In this study, a single nucleotide polymorphism (SNP) typing method, based on 87 SNPs, was developed and compared with PFGE and MLVA. The discriminatory indices of SNP typing, PFGE and MLVA were found to be 0.85, 0.95 and 0.83, respectively. Phylogenetic analysis, using SNP typing as Gold Standard, revealed false homoplasies in the PFGE and MLVA trees. Further, in contrast to the SNP-based tree, the PFGE- and MLVA-based trees did not reveal a positive correlation between root-to-tip distance and the isolation year of strains. Thus PFGE and MLVA do not allow an estimation of the relative age of the selected strains. In conclusion, SNP typing was found to be phylogenetically more informative than PFGE and more discriminative than MLVA. Further, in contrast to PFGE, it is readily standardized allowing interlaboratory comparisons. We applied SNP typing to study strains with a novel allele for the pertussis toxin promoter, *ptxP3*, which have a worldwide distribution and which have replaced the resident *ptxP1* strains in the last 20 years. Previously, we showed that *ptxP3* strains showed increased pertussis toxin expression and that their emergence was associated with increased notification in the Netherlands. SNP typing showed that the *ptxP3* strains isolated in the Americas, Asia, Australia and Europe formed a monophyletic branch which recently diverged from *ptxP1* strains. Two predominant *ptxP3* SNP types were identified which spread worldwide. The widespread use of SNP typing will enhance our understanding of the evolution and global epidemiology of *B. pertussis*.

Introduction

Whooping cough or pertussis, caused by the Gram-negative bacterium *Bordetella pertussis*, is an infection of the human respiratory tract and most severe for young, unvaccinated, infants [1,2]. Vaccination against whooping cough, introduced in most countries in the 1950s and 1960s, has been highly successful in controlling the disease [2]. However, whooping cough resurged in the 1980s and in many developed countries it has become the most prevalent vaccine preventable disease [1,2,3]. Several causes for the increase in whooping cough have been proposed, most prominently waning immunity and pathogen adaptation [2,4,5,6]. Initially, pathogen adaptation involved antigenic divergence between clinical isolates and vaccines strains [7,8,9,10,11,12, 13,14,15]. More recently strains emerged with increased pertussis toxin (Ptx) production which harbored a novel allele for the Ptx promoter, *ptxP3* [16]. In the Netherlands, *ptxP3* strains have largely replaced the resident strains which carry the *ptxP1* allele

and the emergence of *ptxP3* strains was closely associated with the increase in notifications [16]. The *ptxP3* strains have been found in several European countries, Japan and the Americas, suggesting a worldwide distribution [16].

Reliable methods to type bacterial pathogens are important tools in infectious disease control. Strain typing can be used to identify successful clones, to establish genetic relationships, to study pathogen evolution and may serve as an early warning system for epidemics [17]. Mainly three typing methods have been used to study *B. pertussis* populations, Multi-Locus Antigen Sequence Typing (MAST), Pulsed-Field Gel Electrophoresis (PFGE) [18] and Multiple-Locus Variable-Number Tandem Repeat Analysis (MLVA) [17].MAST has been very useful to reveal antigenic divergence between circulating strains and vaccine strains [2,4]. However, the discriminative power of MAST is limited and it is not suitable to establish genetic relationships as it targets genes subjected to immune selection. PFGE is used most widely for typing of *B. pertussis* strains and has a high resolving

power [19]. However, it is laborious and because it is difficult to standardize, interlaboratory comparisons of PFGE patterns are problematic. MLVA results in digital profiles allowing accurate interlaboratory comparisons [13,17,20,21] [www.mlva.net]. Further, compared to PFGE it is much less laborious. Unfortunately, the resolving power of MLVA is lower compared to PFGE. Another issue concerning PFGE and MLVA, addressed in this work, is their ability to reveal genetic relationships between strains. Typing of *B. pertussis* strains based on single nucleotide polymorphisms (SNPs) has not been widely used due to the monomorphic nature of *B. pertussis* [22]. However, recently the genomes of several *B. pertussis* strains have been sequenced, identifying hundreds of novel SNPs and potentially providing the tools for SNP typing [23,24,25]. Based on the genome of seven *B. pertussis* strains, we selected 87 SNPs to type strains from Europe and Africa. The results were compared to those of PFGE and MLVA typing. Further, we used SNP typing to extend our studies on the evolution of the *ptxP3* strains.

Materials and Methods

Strains and culturing

Strains were selected to form a geographically and temporally diverse collection. Strains were isolated in three European countries, Finland (n = 7), the Netherlands (n = 26) and Sweden (n = 34) and in two African countries, Kenya (n = 39) and Senegal (n = 18). Strains were isolated in the period 1971 to 2005 (Table S1). Strain 18323 (B1121, ATCC number 9797) was isolated in the USA in 1947 and used to root the phylogenetic trees. All strains were grown on Bordet Gengou agar supplemented with 15% sheep blood and incubated for 3 days at 35°C.

Allele typing

The alleles for *ptxP* were determined by DNA sequencing as described before [16]. Briefly, bacterial cells were lysed in Tris EDTA buffer (Sigma-Aldrich, Zwijndrecht, Nl, 1.0 M Tris-HCl, containing 0.1 M EDTA) at 95°C for 5 minutes, centrifuged for 1 min at 13000 rpm after which the supernatant was used for PCR.

Single nucleotide polymorphism (SNP) typing

SNPs were identified using the complete genome sequences of the Tohama I strain and of six Dutch *B. pertussis* isolates [23,26]. Based on the seven genome sequences, a set of 120 informative SNPs was selected. For SNP typing, chromosomal DNA was isolated using the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich, Zwijndrecht, Netherlands) following the manufacture's instructions for Gram-negative bacteria. SNPs were determined using the iPLEX® Gold assay (Sequenom Inc, Hamburg, Germany). This assay is based on a multiplex PCR followed by a single primer extension reaction which results in allele-specific differences in mass between primer extension products. The primer extension products were analyzed using MALDI-TOF mass spectrometry. SNPs that were ambiguous in a large number of strains were removed which resulted in a final selection of 87 SNPs which were used in this study (Table S2). For each strain, the 87 SNPs were concatenated to generate a contiguous DNA sequence which was used as character data for clustering (Table S1). The Maximum Parsimony algorithm was applied for clustering using Bionumerics version 6.1 (Applied Maths, Sint-Martens-Latem, Belgium).

Pulsed-field gel electrophoresis (PFGE)

PFGE was performed as described before [18,27]. Most strains were typed by PFGE in previous studies [19,28] while the strains from Africa were analyzed in this study. Band matching was

performed with an optimization of 1% and a position tolerance of 1%. The binary character data was used for clustering with the Maximum Parsimony algorithm in Bionumerics version 6.1 (Applied Maths, Sint-Martens-Latem, Belgium).

Multiple-locus variable-number tandem repeat analysis (MLVA)

For MLVA, the number of repeats in 6 loci (VNTR1, VNTR3A, VNTR3B, VNTR4, VNTR5 and VNTR6) was determined as described previously [17,21]. MLVA profiles were clustered as character data using the Maximum Parsimony algorithm in Bionumerics version 6.1 (Applied Maths, Sint-Martens-Latem, Belgium).

Data analysis

To investigate genetic relationships based on SNPs, PFGE and MLVA, Maximum Parsimony trees were constructed in Bionumerics version 6.1 (Applied Maths, Sint-Martens-Latem, Belgium) and strain 18323 was used to root the trees. For all trees, a bootstrap analysis was performed with 1000 iterations. Correspondence between the SNP, PFGE and MLVA typing with, respectively, the *ptxP* alleles and geographic origin of strains, was determined by calculating the Wallace coefficients using EpiCompare version 1.0 (Ridom GmbH, Wurzburg, Germany) [29]. A high Wallace coefficient, for e.g. SNP typing and *ptxP* alleles, implies that the SNP type accurately predicts the *ptxP* type. In addition, EpiCompare version 1.0 was used to calculate the Discriminatory Index (DI) for each typing method [30]. To explore the relationship between the genetic distances to the root and the isolation year of strains, a linear regression was performed and the R-squared value (R^2) was determined.

Results

Genetic relationships between *B. pertussis* strains based on 87 SNPs

A Maximum Parsimony tree was constructed by concatenating all 87 SNPs of the analyzed strains (Figure 1). *B. pertussis* strain 18323, which harbors the *ptxP4* allele, was used to root the tree as we found that this strain was most closely related to *B. bronchiseptica* in this set of strains (data not shown), consistent with previous studies using multi-locus enzyme electrophoresis and comparative genomic hybridization [31,32,33]. In 125 strains, 14 different SNP types (STs) could be distinguished which resulted in a discriminatory index (DI) of 0.85 (Table 1). Bootstrap values ranged from 63% to 100% (average 89%). Five STs (ST3, ST6, ST7, ST11 and ST12) predominated, representing 14%, 24%, 12%, 11% and 22% of the strains, respectively. Six STs were found only once. The predominance of certain STs may have been caused by phylogenetic discovery bias which leads to collapsing of branches [34,35].

Our collection contained strains from two African (Kenya and Senegal) and three European (Finland, Netherlands and Sweden) countries. A high Wallace coefficient (0.95) (Table 1) was found between ST and the continent of origin of strains. Three STs (ST2-ST4, 21%) were unique for African strains; eight STs (ST7-ST14, 50%) were found only in Europe, while two STs (ST5 and ST6, 28%) were found in both continents (Figure 1). Of the five STs found in Africa, only one (24%) was found in both Kenya and Senegal, suggesting distinct populations. Alternatively, the lack of overlap in STs may be related to the different sampling years, as discussed below.

Since divergence increases over time due to the accumulation of SNPs, a relationship between isolation year and position in the tree is to be expected. Indeed, such a relationship was found, when the number of SNPs relative to the root was plotted against the isolation year of strains ($R^2 = 0.65$, P<0.005) (Figure 2)

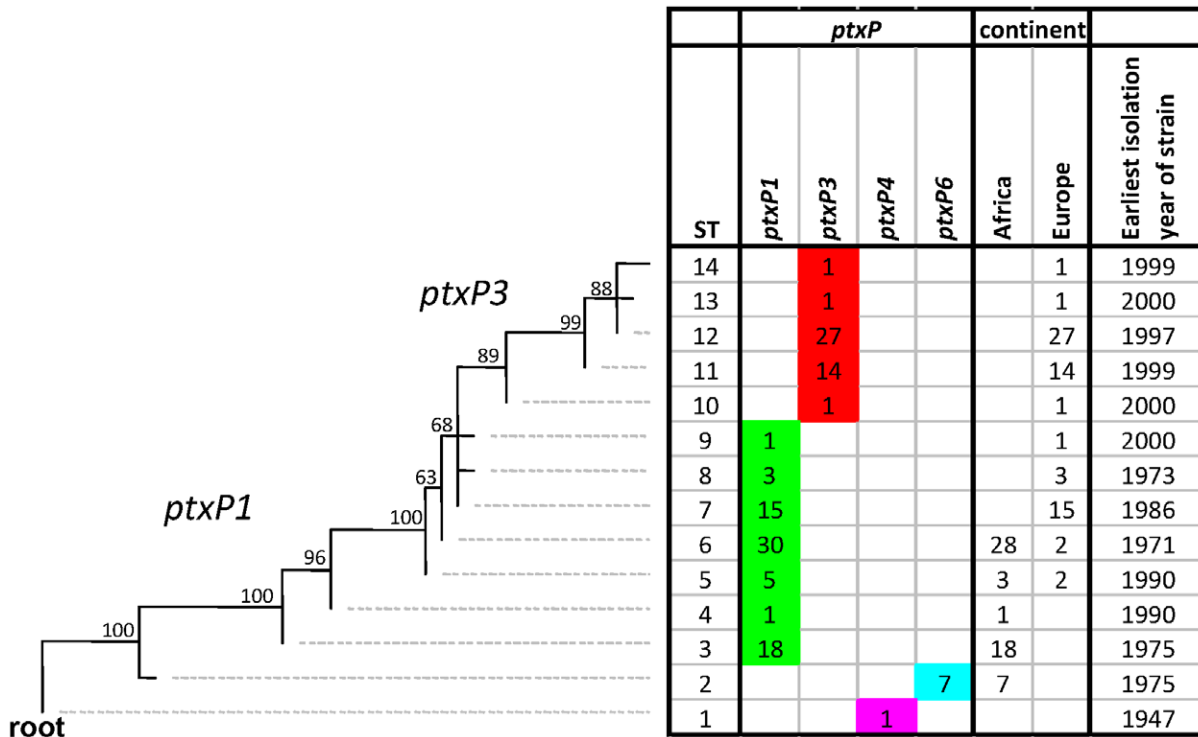| ST	ptxP1	ptxP3	ptxP4	ptxP6	Africa	Europe	Earliest isolation year of strain
14		1				1	1999
13		1				1	2000
12		27				27	1997
11		14				14	1999
10		1				1	2000
9	1					1	2000
8	3					3	1973
7	15					15	1986
6	30				28	2	1971
5	5				3	2	1990
4	1				1		1990
3	18				18		1975
2				7	7		1975
1			1				1947

Figure 1. Maximum Parsimony tree of 125 *B. pertussis* isolates based on 87 SNPs. The *ptxP* alleles, continent of origin and the year of isolation of the oldest isolate are indicated. The numbers refer the number of isolates for each ST. Bootstrap values are indicated in the tree. Abbreviation, ST; SNP type.
doi:10.1371/journal.pone.0020340.g001

A high Wallace coefficient (1.0) (Table 1) was found between ST and *ptxP* allele, as each ST was associated with a single *ptxP* allele. Further, the predominant *ptxP* alleles, *ptxP1* and *ptxP3*, were separated into distinct clusters. The *ptxP6* allele was identified in African strains only. The *ptxP6* strains were placed close to the root suggesting they represent an ancient lineage. No evidence for homoplasy was found for *ptxP* alleles in the SNP-based tree. The tree indicated that the *ptxP3* strains diverged from *ptxP1* strains, as suggested previously [16]. Further, the relative large genetic distance between the *ptxP3* branch and the root suggested that this event occurred recently. Consistent with this, the earliest isolation year for a *ptxP3* strain in the Netherlands was 1988.

Genetic relationships between *B. pertussis* strains based on PFGE

PFGE analysis of 125 strains revealed 37 unique PFGE types which resulted in a DI of 0.95 (Figure 3, Table 1). Bootstrap values

were lower compared to the SNP-based tree and ranged from 0% to 100% (average 14%). Nine PFGE types (SR3, SR10, SR11, SR5, SR23, SR127, AFR1, AFR5 and FINR9) predominated, representing 4%, 5%, 13%, 4%, 4%, 6%, 13%, 6% and 10% of all strains, respectively. Fourteen PFGE types were found only once.

A high Wallace coefficient (0.99) (Table 1) was found between PFGE typing and continent of origin of strains. Ten PFGE types (42%) were specific for Africa, 25 types (53%) were found only in Europe. One PFGE type (4%), SR23, was shared by the two continents but was found only once in Europe (Sweden) [36]. Of the ten African PFGE types, only one, FINR9, (10%) was found in both Kenya and in Senegal which suggested two distinct populations as mentioned above. In contrast to the SNP-based tree, a negative relationship between isolation year and tip-to-root distance was found ($R^2 = 0.29$, P<0.005) (Figure S1).

A high Wallace coefficient (0.99) (Table 1) was found between PFGE type and *ptxP* allele, as each PFGE type was associated with a single *ptxP* allele, except for SR10, which was shared by both *ptxP1* and *ptxP3* strains. The African *ptxP6* lineage was represented by one unique PFGE type, AFR5. In contrast to the SNP-based tree, the *ptxP1* and *ptxP3* strains were not separated in distinct clusters as the *ptxP3* strains were interrupted by *ptxP1* strains, suggesting homoplasy. Further, the *ptxP6* lineage was placed close to the tip and root in the PFGE- and SNP-based trees, respectively, suggesting that the *ptxP6* lineage represented, respectively, a recent and an ancient branch.

Genetic relationships between *B. pertussis* strains based on MLVA

MLVA analysis revealed 21 unique MLVA types which resulted in a DI of 0.83 (Figure 4, Table 1). Bootstrap values were lower

Table 1. Wallace coefficients and Discriminatory indices.[1]

Typing method	Continent of origin	ptxP allele	Discriminatory index
SNP typing	0.95	1.00	0.85 [0.83–0.88]
PFGE	0.99	0.99	0.95 [0.93–0.96]
MLVA	0.92	0.82	0.83 [0.78–0.88]

[1]The Wallace coefficient of the typing method was calculated versus, respectively, continent of origin and *ptxP* allele. The confidence interval for the Discriminatory index has been indicated.
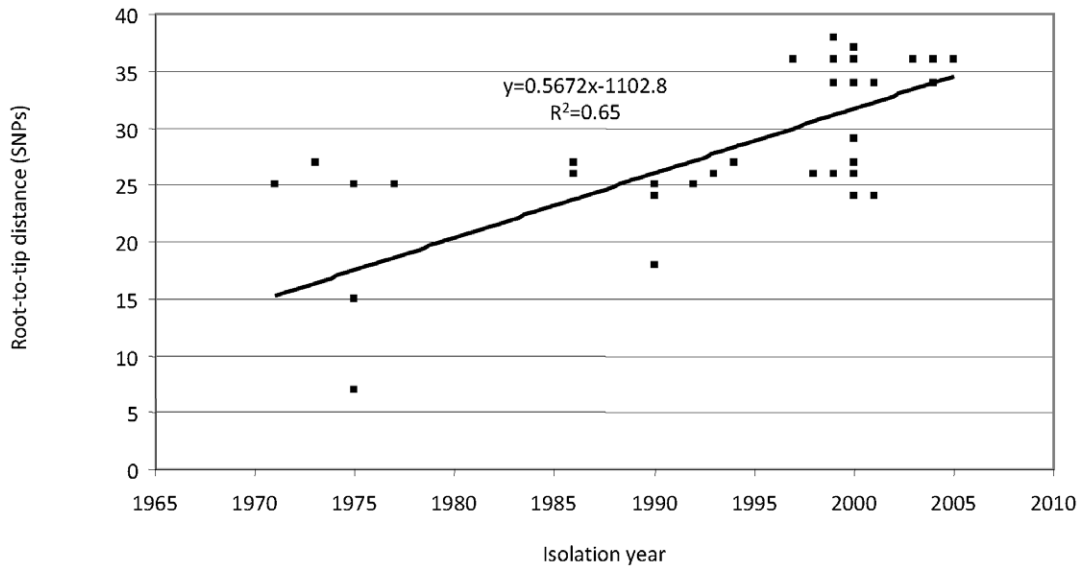doi:10.1371/journal.pone.0020340.t001

Figure 2. Relationship between year of isolation and the-root-to-tip distance in the SNP-based tree. The root-to-tip distance represents the number of SNP differences between a particular clinical isolate and strain 18323, used to root the tree. Linear regression was performed and the trend line and R-squared value (R^2) are indicated. A significant increase of SNPs over time was found ($R^2 = 0.65$, P<0.005).
doi:10.1371/journal.pone.0020340.g002

compared to the SNP-based tree and ranged from 5% to 100% (average 37%). Four MLVA types (MLVA29, MLVA10, MLVA149 and MLVA27) predominated, frequencies were 16%, 9%, 14% and 34% respectively. Eleven MLVA types were found only once.

Although, a high Wallace coefficient (0.92) (Table 1) was found between MLVA type and continent of origin, separation of African strains and European strains was less pronounced compared to the trees based on SNPs. Seven MLVA types (36%) were specific for Africa, ten (42%) were unique for Europe and three MLVA types (21%) were shared by both continents. Of the seven African MLVA types, only one, MLVA96, (5%) was found in both Kenya and in Senegal which suggested two distinct populations similar to trees based on SNPs and PFGE. Similar to PFGE typing, a negative correlation was found between isolation year and distance to the root ($R^2 = 0.31$, P<0.005) (Figure S2).

Compared to the SNP- and PFGE-based trees, a lower Wallace coefficient (0.82) (Table 1) was found between MLVA type and *ptxP* allele. The *ptxP6* lineage was represented by two MLVA types, MLVA49 and MLVA169, which were clustered together. In contrast to the SNP-based tree, but similar to the PFGE-based tree, the *ptxP6* lineage showed a large distance to the root which suggested a recent origin.

Eleven MLVA types (38%) were specific for *ptxP1* strains, five MLVA types (5%) were unique for *ptxP3* strains and two MLVA types, MLVA27 and MLVA29, were shared by both *ptxP1* and *ptxP3* strains (frequencies, 34% and 16% respectively). MLVA27 and MLVA29 are currently the predominant types in Europe and Australia [13,17,21]. The two MLVA types were much better resolved by SNP and PFGE typing, resulting in a subdivision into, respectively, 13 and 15 types for MLVA27 and 7 and 12 for MLVA29.

The *ptxP1* and *ptxP3* strains were well separated in two distinct clusters, except for the MLVA types, MLVA29 and MLVA27 which contained strains with both alleles. The presence of the *ptxP3* allele in the MLVA29 branch, which was mainly comprised of *ptxP1* strains and placed some distance from the main *ptxP3* cluster, suggested homoplasy. Further, as observed for the SNP-

based tree, the MLVA-based tree indicated that most *ptxP3* strains represented a young branch in the tree which diverged from *ptxP1* strains.

Global distribution of the *ptxP3* strain

SNP typing was extended to include 179 *ptxP3* strains from diverse geographic locations; Asia (1.1%), Australia (2.3%), Europe (86%), North America (8.9%) and South America (1.7%) (Figure 5, Table S3). Bootstrap values ranged from 63% to 97% (average 81%). Four new STs were identified (ST15-ST18). A Wallace coefficient of 0.77 was found between ST and continent of origin. Of the nine STs which were specific for the *ptxP3* strains, four (ST13, ST14, ST17 and ST18) were found in only one continent (Europe). The remaining STs were found in strains from two to four continents. Two SNP types, ST11 and ST12, were found to be most prevalent.

Discussion

SNP typing has been used extensively to study the evolution, spread and phylogeography of bacterial pathogens and can be seen as the Gold Standard against which other methods can be compared [37]. *B. pertussis* populations have been mainly studied by PFGE and MLVA. Although these typing methods have been useful for strain discrimination, it is not clear to what extent they reflect genetic relationships. Here we compared the widely used typing systems based on PFGE and MLVA with the newly developed SNP typing.

The strains were isolated in Europe and Africa and the SNP-based tree showed a striking consistency with geographic source. This may reflect geographic isolation. However, the congruence may also reflect the differences in isolation dates between the strains from Europe and Africa (average isolation dates, respectively, 1999 and 1980) or vaccination status of the populations. Most of the European strains (88%) were isolated from Finland, Sweden and the Netherlands in a period when the vaccination coverage was high (>96%) [38]. In Finland and the Netherlands mainly whole cell pertussis vaccines were used, while in Sweden
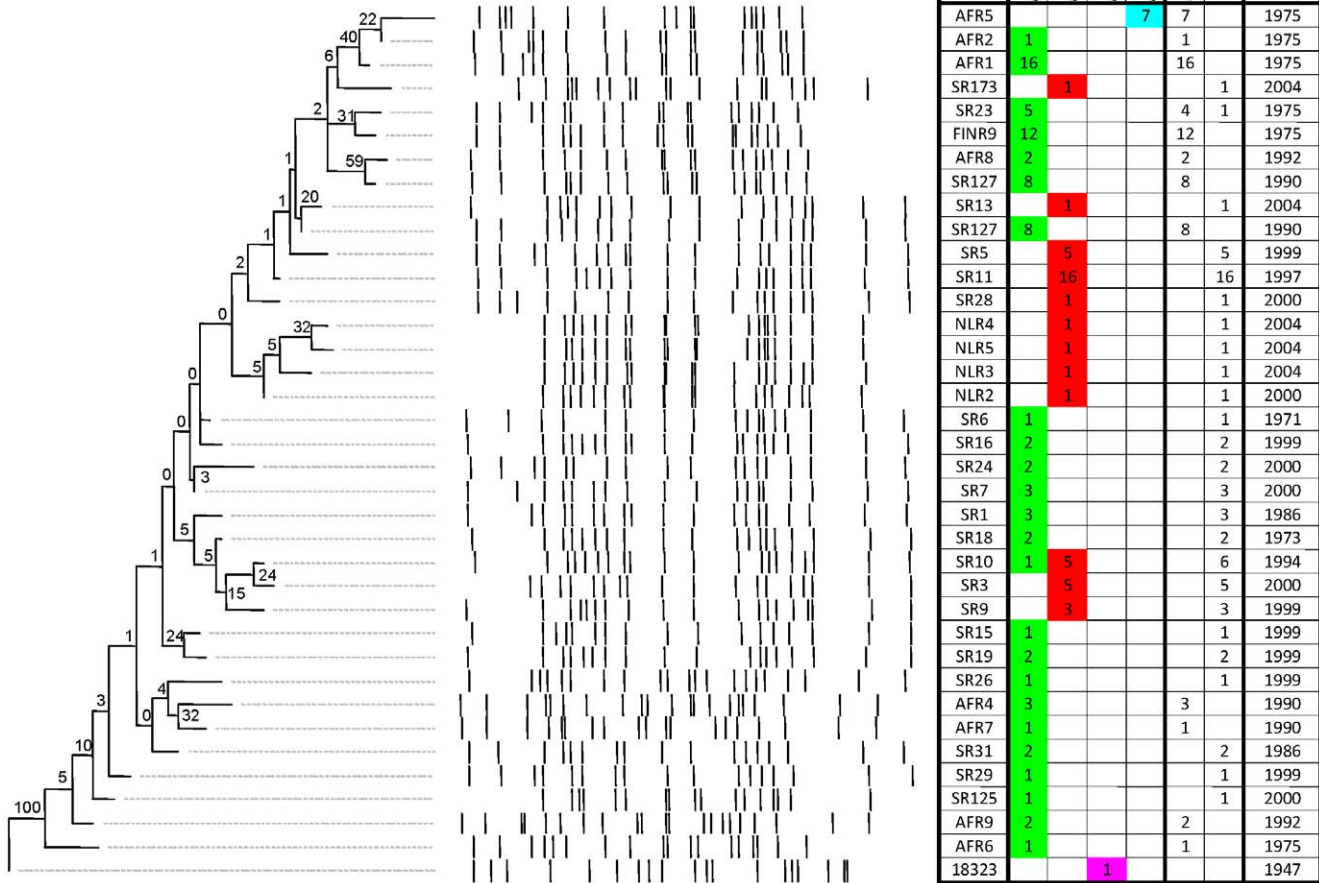
PFGE type	ptxP1	ptxP3	ptxP4	ptxP6	Africa	Europe	Year of isolation oldest isolate
AFR5				7	7		1975
AFR2	1				1		1975
AFR1	16				16		1975
SR173		1				1	2004
SR23	5				4	1	1975
FINR9	12				12		1975
AFR8	2				2		1992
SR127	8				8		1990
SR13		1				1	2004
SR127	8				8		1990
SR5		5				5	1999
SR11		16				16	1997
SR28		1				1	2000
NLR4		1				1	2004
NLR5		1				1	2004
NLR3		1				1	2004
NLR2		1				1	2000
SR6	1					1	1971
SR16	2					2	1999
SR24	2					2	2000
SR7	3					3	2000
SR1	3					3	1986
SR18	2					2	1973
SR10	1	5				6	1994
SR3		5				5	2000
SR9		3				3	1999
SR15	1					1	1999
SR19	2					2	1999
SR26	1					1	1999
AFR4	3				3		1990
AFR7	1				1		1990
SR31	2					2	1986
SR29	1					1	1999
SR125	1					1	2000
AFR9	2				2		1992
AFR6	1				1		1975
18323			1				1947

Figure 3. Maximum Parsimony tree of 125 *B. pertussis* isolates based on PFGE. See legends to Fig. 1 for further details.
doi:10.1371/journal.pone.0020340.g003

acellular vaccines were used. The strains from Kenya were isolated in a period when only approximately 20% of the children had a record of receiving three doses of a pertussis whole cell vaccine, while 10% received only two doses [39]. Strains from Senegal were isolated in 1990 to 1992, three to five years after the introduction of vaccination with a whole cell vaccine (i.e. in 1987) [40]. The differences in vaccination status between Europe and Africa may explain the distinct clustering of European and African strains in the SNP-based tree.

In the strain collection used in this study, one *ptxP* allele, *ptxP6*, was found only in the Kenyan strains (N = 7), isolated in 1975 [41]. Analysis of *ptxP* alleles in a larger collection of strains also revealed the presence of *ptxP6* in two Dutch strains, both isolated in 1977 (data not shown) and it is conceivable that these strains were imported from Africa. Further studies are required to determine if the *ptxP6* allele is characteristic for African strains. No distinct African alleles were found for the virulence factors Ptx, pertactin, Fim2 and Fim3 (data not shown). In conclusion, at present we cannot conclude that the African strains form a distinct population.

Lan and coworkers have used a DNA tiling microarray-based approach to detect SNPs and study the evolution of *B. pertussis*

[25,42]. In contrast to our study, which covered the complete genome, Lan and coworkers targeted 34% (1.4 Mb) of the genome. Using 65 SNPs, 42 STs were identified in 316 isolates [42]. The DI obtained by Lan and coworkers was higher than the DI obtained by us (respectively, 0.94 and 0.85) and this may be due to the fact that their strain collection was more diverse with respect to geographic origin and year of isolation. Similar to our results with *ptxP* alleles, they found that surface protein alleles correlated with evolutionary lineages as revealed by SNP-typing. One branch (cluster I) identified by Lan and coworkers contained strains with a worldwide distribution, and we speculate that this branch contains *ptxP3* strains. The results of Lan and coworkers are consistent with vaccine-drive evolution of *B. pertussis* as first proposed by us [2,6,14].

When we compared the resolving powers of SNP typing, PFGE and MLVA, the highest DI was found for PFGE (DI 0.95) followed by SNP typing (DI 0.85) and MLVA (DI 0.83). The limitations of MLVA were illustrated by the inability to resolve MLVA27 and MLVA29. These MLVA types represented 50% of all analyzed strains. Both types have a worldwide distribution [13,17,20,43] and in particular MLVA27 is associated with the
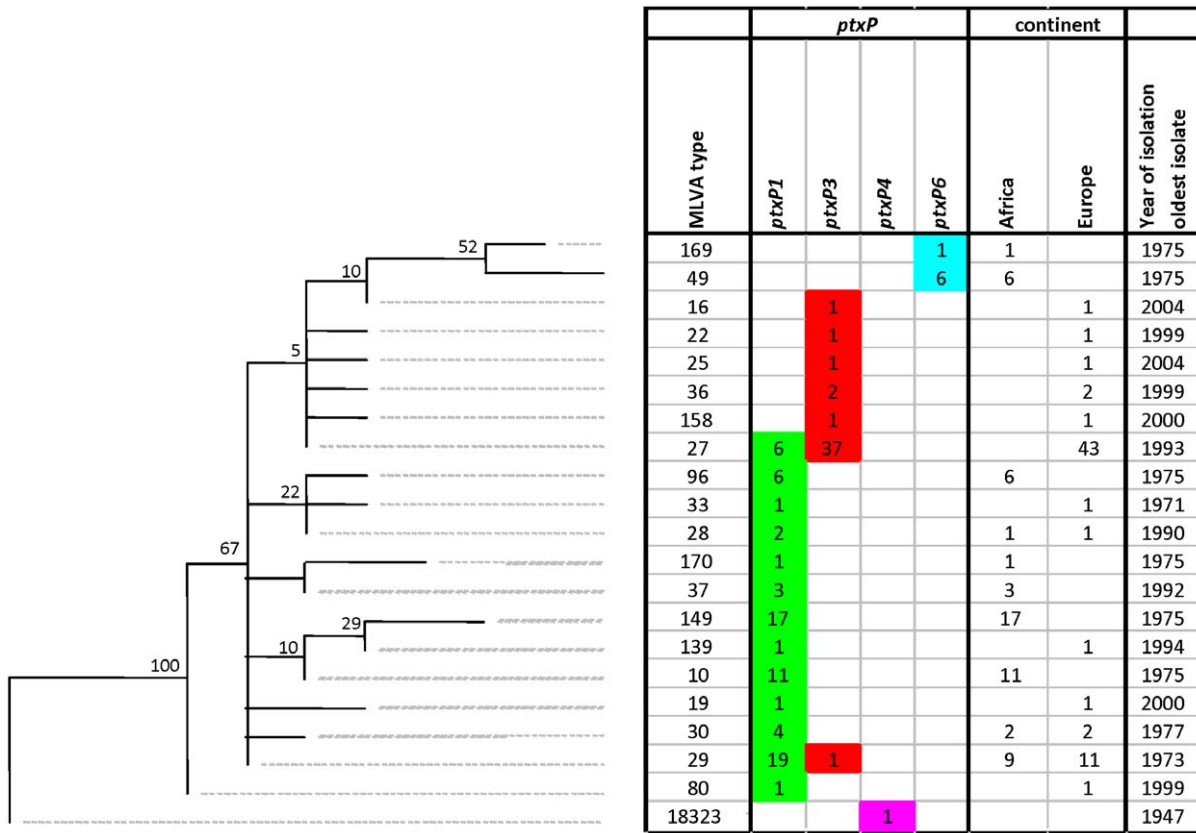
MLVA type	ptxP1	ptxP3	ptxP4	ptxP6	Africa	Europe	Year of isolation oldest isolate
169				1	1		1975
49				6	6		1975
16		1				1	2004
22		1				1	1999
25		1				1	2004
36		2				2	1999
158		1				1	2000
27	6	37				43	1993
96	6				6		1975
33	1					1	1971
28	2				1	1	1990
170	1				1		1975
37	3				3		1992
149	17				17		1975
139	1					1	1994
10	11				11		1975
19	1					1	2000
30	4				2	2	1977
29	19	1			9	11	1973
80	1					1	1999
18323			1				1947

Figure 4. Maximum Parsimony tree of 125 *B. pertussis* isolates based on MLVA. See legends to Fig. 1 for further details.
doi:10.1371/journal.pone.0020340.g004

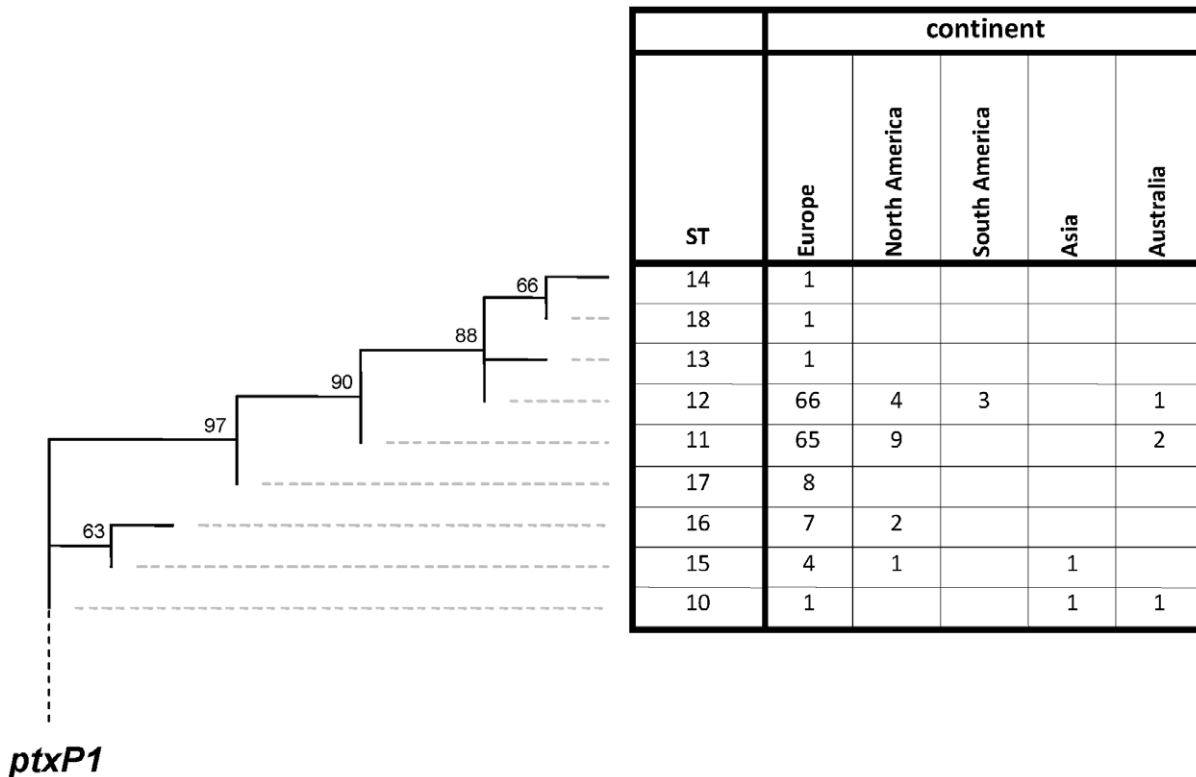| ST	Europe	North America	South America	Asia	Australia
14	1				
18	1				
13	1				
12	66	4	3		1
11	65	9			2
17	8				
16	7	2			
15	4	1		1	
10	1			1	1

ptxP1

Figure 5. Maximum Parsimony tree of 179 *ptxP3* strains found worldwide based on 87 SNPs. See legends to Fig. 1 for further details.
doi:10.1371/journal.pone.0020340.g005

ptxP3 allele [9,13,21]. The two MLVA types were resolved into, respectively, six and five types by SNP-typing and 15 and 12 types by PFGE. In contrast to the SNP-based tree, the PFGE- and MLVA-based trees revealed several examples of homoplasies with respect to *ptxP* alleles. Homoplasy with respect to *ptxP3* was also suggested by a previous study using PFGE [43]. The SNP analysis indicated that these are false homoplasies, probably caused by convergence of PFGE and MLVA types. Recombination between insertion sequences and slipped strand mispairing, play an important role in generating diversity in PFGE and MLVA types, respectively. Both processes are reversible and may thus lead to convergence and false genetic relationships. A linear regression analysis of the root-to-tip distance against the year of isolation showed a positive correlation for the SNP-based tree, reflecting the gradual accumulation of point mutations. Interestingly, a negative correlation was found for both PFGE and MLVA, implying that estimation of the relative age of strains is not possible with these methods. Overall, our comparison shows that in particular typing based on PFGE did not reflect true genetic relationships and is therefore less suitable for evolutionary studies of *B. pertussis*. However, in view of its high resolving power PFGE remains a useful tool for strain discrimination. Results obtained with MLVA are more consistent with SNP typing. However, the resolving power of MLVA is rather limited. It remains useful as a high throughput, highly portable, typing method, however.

We used SNP typing to study the genetic relationships between strains of the *ptxP3* lineage which have been associated with the increase in pertussis in the Netherlands [16]. The SNP-based tree suggested that *ptxP3* strains, isolated in Asia, the Americas, Australia and Europe, formed a monophyletic group which evolved from *ptxP1* strains. The tree is consistent with the assumption that a single clone spread globally and diversified. In particular two *ptxP3* STs (ST11 and ST12) were found to predominate, suggesting fitness differences between *ptxP3* strains. It should be noted however that the predominance of STs may have been caused by phylogenetic discovery bias which leads to collapsing of branches [34,35]. The placement of the *ptxP3* branch at the tip of the tree suggested that *ptxP3* strains evolved relatively recently, possibly after the introduction of vaccination. Consistent with this assumption, *ptxP3* strains were first detected in the Netherlands in 1988 and subsequently increased in prevalence to ≥90% in 2002. The absence of *ptxP3* strains from populations in Africa with low vaccine coverage (Kenya), or where vaccination was introduced recently (Senegal), also suggests that *ptxP3* strains evolved due to the introduction of vaccination.

In summary, the widespread use of SNP typing will enhance our understanding of the evolution and global epidemiology of *B. pertussis*.

Supporting Information

Figure S1 Relationship between year of isolation and root-to-tip distance in the PFGE-based tree. The root-to-tip distance between a particular isolate and strain 18323, used to root the tree, was based on the genetic distance. Linear regression was performed and the trend line and R-squared value (R^2) are indicated. A negative correlation was found between isolation year and distance to the root ($R^2 = 0.29$, P<0.005).
(TIFF)

Figure S2 Relationship between year of isolation and root-to-tip distance in the MLVA-based tree. The root to tip distance between a particular isolate and strain 18323, used to root the tree, was based on the genetic distance. Linear regression was performed and the trend line and R-squared value (R^2) are indicated. A negative correlation was found between isolation year and distance to the root ($R^2 = 0.31$, P<0.005).
(TIFF)

Table S1 Strains isolated from 1971 to 2005.
(XLS)

Table S2 87 ambiguous SNPs.
(XLS)

Table S3 SNP typing extended to include 179 ptxP3 strains from diverse geographic locations.
(XLS)

Author Contributions

Conceived and designed the experiments: MvG MJB HGJvdH TK QH JM AA HOH FRM. Performed the experiments: MvG MJB HGJvdH KJH TK AA. Analyzed the data: MvG MJB HGJvdH KJ FRM. Contributed reagents/materials/analysis tools: KJ. Wrote the paper: MvG PWH FRM.

References

1. He Q, Mertsola J (2008) Factors contributing to pertussis resurgence. Future Microbiol 3: 329–339.
2. Mooi FR (2010) Bordetella pertussis and vaccination: the persistence of a genetically monomorphic pathogen. InfectGenetEvol 10: 36–49.
3. de Greeff SC, de Melker HE, van Gageldonk PG, Schellekens JF, van der Klis FR, et al. (2010) Seroprevalence of Pertussis in the Netherlands: Evidence for Increased Circulation of Bordetella pertussis. PLoS One 5: e14183.
4. van Boven M, Mooi FR, Schellekens JF, de Melker HE, Kretzschmar M (2005) Pathogen adaptation under imperfect vaccination: implications for pertussis. Proc Biol Sci 272: 1617–1624.
5. Kallonen T, He Q (2009) Bordetella pertussis strain variation and evolution postvaccination. ExpertRevVaccines 8: 863–875.
6. Mooi FR, van Loo IH, King AJ (2001) Adaptation of Bordetella pertussis to vaccination: a cause for its reemergence? EmergInfectDis 7: 526–528.
7. Borisova O, Kombarova SY, Zakharova NS, van Gent M, Aleshkin VA, et al. (2007) Antigenic divergence between Bordetella pertussis clinical isolates from Moscow, Russia, and vaccine strains. ClinVaccine Immunol 14: 234–238.
8. Fingermann M, Fernandez J, Sisti F, Rodriguez ME, Gatti B, et al. (2006) Differences of circulating Bordetella pertussis population in Argentina from the strain used in vaccine production. Vaccine 24: 3513–3521.
9. Fry NK, Neal S, Harrison TG, Miller E, Matthews R, et al. (2001) Genotypic variation in the Bordetella pertussis virulence factors pertactin and pertussis toxin in historical and recent clinical isolates in the United Kingdom. InfectImmun 69: 5520–5528.
10. Gzyl A, Augustynowicz E, van Loo I, Slusarczyk J (2001) Temporal nucleotide changes in pertactin and pertussis toxin genes in Bordetella pertussis strains isolated from clinical cases in Poland. Vaccine 20: 299–303.
11. Hallander HO, Advani A, Donnelly D, Gustafsson L, Carlsson RM (2005) Shifts of Bordetella pertussis variants in Sweden from 1970 to 2003, during three periods marked by different vaccination programs. JClinMicrobiol 43: 2856–2865.
12. Kodama A, Kamachi K, Horiuchi Y, Konda T, Arakawa Y (2004) Antigenic divergence suggested by correlation between antigenic variation and pulsed-field gel electrophoresis profiles of Bordetella pertussis isolates in Japan. J Clin Microbiol 42: 5453–5457.
13. Kurniawan J, Maharjan RP, Chan WF, Reeves PR, Sintchenko V, et al. (2010) Bordetella pertussis clones identified by multilocus variable-number tandem-repeat analysis. EmergInfectDis 16: 297–300.
14. Mooi FR, van Oirschot H, Heuvelman K, van der Heide HG, Gaastra W, et al. (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. InfectImmun 66: 670–675.
15. Yao SM, Lin YC, Chou CY, Chen YY, Hsiao MJ, et al. (2005) Antigenic divergence of Bordetella pertussis isolates in Taiwan. J Clin Microbiol 43: 5457–5461.

16. Mooi FR, van Loo IH, van Gent M, He Q, Bart MJ, et al. (2009) Bordetella pertussis strains with increased toxin production associated with pertussis resurgence. EmergInfectDis 15: 1206–1213.

17. van Gent M, de Greeff SC, van der Heide HG, Mooi FR (2009) An investigation into the cause of the 1983 whooping cough epidemic in the Netherlands. Vaccine 27: 1898–1903.

18. Mooi FR, Hallander H, Wirsing von Konig CH, Hoet B, Guiso N (2000) Epidemiological typing of Bordetella pertussis isolates: recommendations for a standard methodology. EurJClinMicrobiolInfectDis 19: 174–181.

19. Advani A, Donnelly D, Hallander H (2004) Reference system for characterization of Bordetella pertussis pulsed-field gel electrophoresis profiles. JClinMicrobiol 42: 2890–2897.

20. Litt DJ, Neal SE, Fry NK (2009) Changes in genetic diversity of the Bordetella pertussis population in the United Kingdom between 1920 and 2006 reflect vaccination coverage and emergence of a single dominant clonal type. J Clin Microbiol 47: 680–688.

21. 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. JBacteriol 186: 5496–5505.

22. Diavatopoulos DA, Cummings CA, Schouls LM, Brinig MM, Relman DA, et al. (2005) Bordetella pertussis, the causative agent of whooping cough, evolved from a distinct, human-associated lineage of B. bronchiseptica. PLoS Pathog 1: e45.

23. Bart MJ, van Gent M, van der Heide HG, Boekhorst J, Hermans P, et al. (2010) Comparative genomics of prevaccination and modern Bordetella pertussis strains. BMCGenomics 11: 627.

24. Bouchez V, Caro V, Levillain E, Guigon G, Guiso N (2008) Genomic content of Bordetella pertussis clinical isolates circulating in areas of intensive children vaccination. PLoS One 3: e2437.

25. Maharjan RP, Gu C, Reeves PR, Sintchenko V, Gilbert GL, et al. (2008) Genome-wide analysis of single nucleotide polymorphisms in Bordetella pertussis using comparative genomic sequencing. Res Microbiol 159: 602–608.

26. Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, et al. (2003) Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. NatGenet 35: 32–40.

27. Advani A, Gustafsson L, Carlsson RM, Donnelly D, Hallander HO (2007) Clinical outcome of pertussis in Sweden: association with pulsed-field gel electrophoresis profiles and serotype. APMIS 115: 736–742.

28. Elomaa A, He Q, Minh NN, Mertsola J (2009) Pertussis before and after the introduction of acellular pertussis vaccines in Finland. Vaccine 27: 5443–5449.

29. Carrico JA, Silva-Costa C, Melo-Cristino J, Pinto FR, de Lencastre H, et al. (2006) Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant Streptococcus pyogenes. JClinMicrobiol 44: 2524–2532.

30. Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. JClinMicrobiol 26: 2465–2466.

31. Cummings CA, Brinig MM, Lepp PW, van de Pas S, Relman DA (2004) Bordetella species are distinguished by patterns of substantial gene loss and host adaptation. J Bacteriol 186: 1484–1492.

32. Musser JM, Hewlett EL, Peppler MS, Selander RK (1986) Genetic diversity and relationships in populations of Bordetella spp. JBacteriol 166: 230–237.

33. van der Zee A, Mooi F, Van Embden J, Musser J (1997) Molecular evolution and host adaptation of Bordetella spp.: phylogenetic analysis using multilocus enzyme electrophoresis and typing with three insertion sequences. J Bacteriol 179: 6609–6617.

34. Pearson T, Busch JD, Ravel J, Read TD, Rhoton SD, et al. (2004) Phylogenetic discovery bias in Bacillus anthracis using single-nucleotide polymorphisms from whole-genome sequencing. ProcNatlAcadSciUSA 101: 13536–13541.

35. Pearson T, Okinaka RT, Foster JT, Keim P (2009) Phylogenetic understanding of clonal populations in an era of whole genome sequencing. InfectGenetEvol 9: 1010–1019.

36. Elomaa A, Advani A, Donnelly D, Antila M, Mertsola J, et al. (2007) Population dynamics of Bordetella pertussis in Finland and Sweden, neighbouring countries with different vaccination histories. Vaccine 25: 918–926.

37. Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu Rev Microbiol 62: 53–70.

38. van Amersfoorth SC, Schouls LM, van der Heide HG, Advani A, Hallander HO, et al. (2005) Analysis of Bordetella pertussis populations in European countries with different vaccination policies. JClinMicrobiol 43: 2837–2843.

39. Voorhoeve AM, Muller AS, Schulpen TW, t' Mannetje W, van Rens M (1978) Machakos project studies. Agents affecting health of mother and child in a rural area of Kenya. IV: The epidemiology of pertussis. Trop Geogr Med 30: 125–139.

40. Preziosi MP, Yam A, Wassilak SG, Chabirand L, Simaga A, et al. (2002) Epidemiology of pertussis in a West African community before and after introduction of a widespread vaccination program. Am J Epidemiol 155: 891–896.

41. Patel S, Schoone G, Ligthart GS, Gikken H, Preston NW (1978) Machakos project studies. Agents affecting health of mother and child in a rural area of Kenya. V. Pertussis sentypes in Kenyan children 1974–1975. TropGeogrMed 30: 141–146.

42. Octavia S, Maharjan RP, Sintchenko V, Stevenson G, Reeves PR, et al. (2010) Insight into evolution of Bordetella pertussis from comparative genomic analysis: evidence of vaccine-driven selection. MolBiolEvol.

43. 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. JMicrobiolMethods 78: 297–301.