Insight into Evolution of <i>Bordetella pertussis</i> from Comparative Genomic Analysis: Evidence of Vaccine-Driven Selection

Sophie Octavia,†1 Ram P. Maharjan,†1 Vitali Sintchenko,2 Gordon Stevenson,3 Peter R. Reeves,3 Gwendolyn L. Gilbert,2 and Ruiting Lan*,1

1School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales, Australia
2Centre for Infectious Diseases and Microbiology, the University of Sydney, Westmead Hospital, Westmead, New South Wales, Australia
3School of Molecular Bioscience, University of Sydney, Sydney, New South Wales, Australia
†These authors contributed equally to this work.

*Corresponding author: E-mail: r.lan@unsw.edu.au.

Associate editor: Daniel Falush

Abstract

Despite high vaccine coverage, pertussis incidence has increased substantially in recent years in many countries. A significant factor that may be contributing to this increase is adaptation to the vaccine by <i>Bordetella pertussis</i>, the causative agent of pertussis. In this study, we first assessed the genetic diversity of <i>B. pertussis</i> by microarray-based comparative genome sequencing of 10 isolates representing diverse genotypes and different years of isolation. We discovered 171 single nucleotide polymorphisms (SNPs) in a total of 1.4 Mb genome analyzed. The frequency of base changes was estimated as one per 32 kb per isolate, confirming that <i>B. pertussis</i> is one of the least variable bacterial pathogens. We then analyzed an international collection of 316 <i>B. pertussis</i> isolates using a subset of 65 of the SNPs and identified 42 distinct SNP profiles (SPs). Phylogenetic analysis grouped the SPs into six clusters. The majority of recent isolates belonged to clusters I–IV and were descendants of a single prevaccine lineage. Cluster I appeared to be a major clone with a worldwide distribution. Typing of genes encoding acellular vaccine (ACV) antigens, ptxA, prn, fhaB, fim2, and fim3 revealed the emergence and increasing incidence of non-ACV alleles occurring in clusters I and IV, which may have been driven by ACV immune selection. Our findings suggest that <i>B. pertussis</i>, despite its high population homogeneity, is evolving in response to vaccination pressure with recent expansion of clones carrying variants of genes encoding ACV antigens.

Key words: <i>Bordetella pertussis</i>, evolution, single nucleotide polymorphism, comparative genome sequencing.

Introduction

Pertussis, or whooping cough, is a highly contagious disease caused by an obligate human pathogen <i>Bordetella pertussis</i>. The disease can be life threatening in infants and young children but is relatively mild in adults and adolescents. However, infected adults and adolescents serve as a reservoir for transmission to unvaccinated infants (Celentano et al. 2005; Tan et al. 2005). After the apparent success of mass vaccination in the 1950s, a resurgence of pertussis has been observed across the globe (Mooi et al. 2001; Poynten et al. 2004; Eloama et al. 2007; Halperin 2007). According to a recent estimate, over 30 million cases of pertussis occur worldwide each year causing 400,000 deaths, mostly in developing countries (World Health Organization 1999). A number of potential explanations for this have been proposed, including better recognition and laboratory diagnosis (and therefore notification) of pertussis in older age groups. However, there is also accumulating evidence that the emergence of new antigenic variants of <i>B. pertussis</i> may reduce the protective efficacy of current vaccines (Mooi et al. 2001; Poynten et al. 2004; Schouls et al. 2004; Byrne and Slack 2006; Litt et al. 2009; Mooi 2009).

<i>Bordetella pertussis</i> has an array of virulence determinants, such as pertussis toxin (Ptx); adenylate cyclase-hemolysin; and adhesins, such as filamentous hemagglutinin (Fha), fimbriae (Fim), and pertactin (Prn) (Mattoo and Cherry 2005). It has been proposed that a major factor in the resurgence of pertussis is the genetic adaptation of circulating strains to immune pressure from vaccination (Mooi et al. 2001). In particular, the introduction of acellular vaccines (ACVs) reduced the number of antigens in whole-cell vaccines (WCV) to three or five antigens (Ptx, Prn, Fha, Fim2, and Fim3). ACVs, first introduced in Japan in the 1980s, are now widely used (Tan et al. 2005), and variation in genes encoding these antigenic components has been reported in many countries, including Australia, France, the Netherlands, United States, United kingdom, and Poland (Tan et al. 2005). However, the association between strain variation and vaccination remains controversial. For example, Godfroid et al. (2005) believe that the current data do not support a direct association between strain variation and vaccine selection pressure except in the case of fimbrial antigens (Fim2 and Fim3). An understanding of the global population structure...
and evolutionary history of circulating B. pertussis is therefore critical for deciphering whether and how this pathogen has adapted to immune pressure due to vaccine use. In this study, we describe the evolutionary sequence of the appearance of nonvaccine alleles of the ACV antigen genes and associated expansion of clones that carry them, which provides new evidence to inform the debate.

Multilocus sequence typing has demonstrated that B. pertussis is a recently derived clone of B. bronchiseptica with very low genetic diversity (Diavatopoulos et al. 2005). We recently applied multilocus variable number tandem repeat (VNTR) analysis (MLVA) to the study of the population structure of B. pertussis (Kurniawan et al. 2010). Representative isolates from Finland, Canada, United States, and Japan and a large collection of Australian isolates were analyzed. Six predominant MLVA types (MTs) were found (Kurniawan et al. 2010). Two of them (MT27 and MT29) were widely distributed across the globe, whereas four predominated in specific countries. Several MTs have persisted over decades, including MT27 and MT29, which have circulated for at least half a century. MT27 has also been identified as one of the dominant clones to have emerged in United Kingdom (Litt et al. 2009). However, VNTR loci may change rapidly, making MLVA unreliable for inferring evolutionary relationships (Octavia and Lan 2009). In addition, as other Bordetella spp. cannot be used as out-groups due to their highly divergent MTs, the inferred phylogeny is unrooted and the direction of change cannot be determined (Kurniawan et al. 2010).

Single nucleotide polymorphisms (SNPs) are the most common form of genetic variation that can be quantified, and genome-wide SNPs have been valuable markers for evolutionary study of genetically homogeneous pathogens (Achtman 2008). However, SNPs are rare in B. pertussis genomes and their identification remains difficult. In a previous study, we detected only 70 SNPs when 1.4 Mb of the 4.09 Mb genome of a recent Australian isolate was compared with the sequenced Tohama I genome using microarray-based comparative genome sequencing (CGS) (Maharjan et al. 2008). In order to gain further insight into the evolution and population structure of B. pertussis and possible immune selection, we performed CGS analysis of an additional 10 strains and selected 65 SNPs to type 316 clinical B. pertussis isolates collected over four decades from 11 countries in four continents.

Materials and Methods

Bacterial Strains and Genomic DNA Preparation

Bacterial isolates were grown on Regan–Lowe medium containing charcoal agar (Oxoid, United Kingdom, Hampshire), 10% horse blood, and Bordetella selective supplement cephalexin (Oxoid) at 37 °C for 3–5 days in a humidified incubator. Genomic DNA was extracted and purified from plate cultures using the phenol–chloroform method as described by Octavia and Lan (2006) and stored at 4 °C.

Details of the strains used in this study are given in supplementary table S3 (Supplementary Material online). The majority were recent isolates including 208 from Australia (1970–2008), 49 from Japan (1989–2007) of which 36 are representatives from the 48 pulsed field gel electrophoresis (PFGE) types identified previously by Kodama et al. (2004), 15 from Finland (1991–2006), 11 from Hong Kong (2002–2006), 8 from France (1993–2007), 12 from United States (2002–2006), 8 from Canada (1994–2005), and 1 each from the Netherlands (1992), Italy (1994), China (1957), United Kingdom (1920), and Mexico (year of isolation unknown). The Canada, Finland, and United States isolates were representatives of the most frequent PFGE types from the respective countries and therefore represent the prevalent strains currently circulating in these countries. The genome sequence of the Tohama I strain was also used as needed. Nine strains, including Tohama I, were isolated before 1960, comprising one each from China, France, Japan, and United Kingdom and five from United States.

Comparative Genomic Analysis and SNP Discovery

The strains used for CGS analysis are listed in table 1. The majority of strains were from our previous study using MLVA (Kurniawan et al. 2010), with isolates selected to represent the predominant MLVA clones of recent years (MT27, MT29, MT64, MT70, MT186, and MT203). We selected two isolates each from the two international MTs (MT27 and MT29), one from Australia, and one from another country. FR0743 (MT109) from France was selected as it represents a predominant PFGE type and by MLVA is more distantly related to other types. Two strains from the 1920s and 1940s, respectively, were also selected to represent diversity in the prevaccine era.

CGS was performed by NimbleGen Systems, covering 1,497 chromosomal regions for a total of 1,391,550 bp or 34% of the Tohama I genome, including 1,012 backbone, B. pertussis-specific and virulence-related genes, 217 pseudogenes, and 268 intergenic (IG) regions. All genes and other regions used for SNP discovery were as described previously (Maharjan et al. 2008).

Real-Time PCR and High Resolution Melt for Typing

There were only two alleles at each of the sites for the SNPs selected, so each SNP required only three primers for hairpin primer real-time polymerase chain reaction (PCR) (HP-RT-PCR); hairpin primer with one each corresponding to the two alleles and one common primer. Typing of SNPs by HP-RT-PCR was as done previously (Maharjan et al. 2008). Typing of antigen genes, which also have length polymorphisms, which required high resolution melt analysis, was done using the scheme described in Chan et al. (2009).

Bioinformatics

A maximum parsimony tree was constructed by heuristic search using PAUP version 4.0 (Swofford 1998). The bootstrap analysis with 1,000 replications was also generated using PAUP. A phylogenetic network based on the neighbor-net algorithm was done using the SplitsTree4 program (Huson 1998). The relationships of changes in the antigen genes...
leading to the differences in antigen-gene profiles (APs) were visualized using eBURST analysis (Feil et al. 2004).

Results

SNP Discovery by Comparative Genomics of *B. pertussis*

To investigate the extent of SNPs in *B. pertussis*, 10 isolates were selected for NimbleGen CGS microarray analysis as described previously (Maharjan et al. 2008), covering 1,497 chromosomal regions of 1,391,550 bp or 34% of the Tohama I genome. CGS discovered 58–111 SNPs in each of nine isolates (excluding *B. pertussis* strain 18323) (table 1) and 585 SNPs for strain 18323. We validated 159 SNPs in the nine isolates, with 45–66 genuine SNPs in each (table 1) and an average false positive rate of 31%. For strain 18323, we validated 60 of 71 SNPs randomly selected from the 585 detected by CGS, which translated into approximately 490 genuine SNPs in strain 18323. This strain is significantly different from other strains and was excluded from analysis of variation. A full list of validated SNPs can be found in supplementary table S1 (Supplementary Material online).

Most of the polymorphic genes contained only a single SNP, but multiple SNPs were detected in seven genes. BP2670, BP2749, BP2946, BP2988, BP3113, and BP3783 each contained two SNPs and BP3092 had three SNPs. The two SNPs in BP2670, BP2749, and BP2946 were present only in L517, BP125, and L900, respectively, and are likely to be due to recombinational events. Of the two SNPs in BP3783 that are present in L900, one is also present in multiple strains. Thus, the two SNPs in L900 are more likely to be due to independent mutational events. The two SNPs in BP2988 and BP3113 are present in nine and eight *B. pertussis* strains, respectively, and could have arisen quite early as both are absent in Tohama I and the 1920s strain NCTC10901. All three of the SNPs in BP3092 are present in L500, two in 02M7492, and one in seven other strains, and all are likely to be due to independent mutational events.

We further looked at whether the genes with multiple SNPs were under selection. BP3783 encodes Ptx subunit A and is a gene known to be under selection as discussed below. One of the two SNPs is a nonsynonymous (ns) SNP, which is a previously known SNP and is in our scheme for typing of antigen genes (Chan et al. 2009). This SNP separates *ptxA1* from the other *ptxA* alleles, and the significance of *ptxA* allele changes is discussed below. Another gene, BP2749 encoding a trifunctional transcriptional regulator, has nsSNPs only, both of which are nonconservative changes. For the remaining genes, two, BP2946 and BP3092, have one synonymous SNP (sSNP) and one nsSNP, whereas three including a pseudogene, BP2670, BP2988, and BP3113, have only sSNPs.

The level of variation between isolates analyzed by CGS is shown in table 1. One hundred and seventy-one SNPs were identified in total (supplementary table S1, Supplementary Material online). One hundred and one were located in backbone genes, 16 in virulence-associated genes, 4 in *B. pertussis*-specific genes, 28 in pseudogenes, and 12 in IG regions (table 1). The frequency of SNPs in different classes of genes and different regions is similar to those found previously in L517 (Maharjan et al. 2008). Eighty-three of the 131 SNPs (63%) in coding regions were nsSNPs: 69 in backbone genes, 1 in a *B. pertussis*-specific gene, and 13 in virulence-associated genes (supplementary tables S1 and S2, Supplementary Material online). The frequency of SNPs in pairwise comparisons ranged from 1 per 16 kb (between NCTC10901 and L900) to 1 per 155 kb (between BP312 and FR0743) and averaged 1 per 32 kb. The combined SNP frequency for all 10 strains is 1 per 8 kb.

**Worldwide Genetic Diversity of *B. pertussis***

We first used 184 SNPs (37 from 18323 and 147 of the 159 SNPs from the other CGS isolates (12 SNPs with less reliable real-time PCR assay results were excluded)) to screen a panel of 29 isolates, representing 18 MTs with multiple isolates from the major types. A phylogenetic tree of the 29 isolates was generated from the SNP data (supplementary fig. S1, Supplementary Material online) and used to select SNPs for comprehensive typing. We selected at least one SNP each for internal branches. For terminal branches,
the SNPs selected covered all CGS isolates branches except for FR0743 and 18323. There is no SNP in the FR0743 terminal branch, whereas all 37 SNPs from 18323 that were used in the analysis are not found in any other isolate. Because 18323 can be differentiated from the other isolates, we did not include any SNPs on this branch as it only adds to the branch length of 18323 with no further resolution achieved. The 65 SNPs, which covered all internal nodes and most terminal branches were then used to type 316 B. pertussis isolates. The SNPs differentiated the isolates into 42 SNP profiles (SPs) except that one isolate could not been fully typed (supplementary table S3, Supplementary Material online). The number of SNP differences in pairwise comparisons of the 42 SPs ranged from 1 to 33 SNPs, with an average of 11.4. Ten SPs were represented by single isolates, whereas the remaining 32 were shared by multiple isolates. SP14 and SP37 were the most common SPs, represented by 41 and 34 isolates, respectively.

Twelve SPs contained isolates originating from two or more countries, and four SPs were found in five or more countries, being SP13 in six countries, SP14 in seven countries, SP11 and SP18 each in five countries. On the other hand, SP26 (14 isolates) and SP30 (25 isolates) were found only in Japan and Australia, respectively.

**Phylogenetic Relationships of the SPs**

A heuristic search found 18 most parsimonious trees of equal length. The trees differed in the placement of SP8, SP9, SP23, and SP29, and one of the trees closest to the consensus tree is shown in Figure 1. We also constructed a neighbor-net tree, which revealed the presence of network structures (supplementary fig. S2, Supplementary Material online). Both trees had six well-defined clusters with the same strain distribution. Ten SNPs were identified as having undergone reverse or parallel changes, and when these were excluded, the network structure on the neighbor-net tree disappeared (supplementary fig. S3, Supplementary Material online), but the six clusters were not changed. Maximum parsimony analysis without the 10 conflicting SNPs generated five trees of equal length with homoplasy index of zero (supplementary fig. S3, Supplementary Material online), which again retain the six clusters, all of which suggest that the tree in Figure 1 represents the true relationships of the six clusters. Loss of resolution of branches was observed for some SPs, but the major branching patterns remained unchanged when the 10 SNPs were excluded.

The maximum parsimony tree (fig. 1) was rooted using both *Bordetella parapertussis* and *B. bronchiseptica*, which suggested that most of the isolates diverged recently. The nine isolates from the 1920s to 1950s are from five countries (China, France, Japan, United Kingdom, and Unites States) and are distributed among six SPs (SP25 and SP32–SP36) that were located close to the root of the tree. The highly divergent strain 18323 is in SP32 with the one other isolate being from France recovered in 1993. Analysis of the 60 confirmed CGS SNPs showed that 18323 carried fewer ancestral alleles (27 SNPs) than Tohama I (32 SNPs) indicating that the placement is correct. This strain appears to have a higher rate of accumulating mutations, perhaps due to mutation in *mutS*, one of the genes responsible for mismatch repair systems as three nsSNPs, at locations 805, 1039, and 1213, were found in the CGS analysis.

The six clusters discussed above each had at least one SNP supporting the grouping. Cluster I was supported by SNP 47; cluster II by SNP 16; cluster III by SNPs 7 and 56; cluster IV by SNPs 33 and 35; cluster V by SNPs 15, 36, and 63; and cluster VI by SNPs 42, 44, and 60. The six clusters had different geographic distributions (fig. 1). Cluster I is spread worldwide with 74 isolates obtained from seven countries between 1994 and 2008. Clusters II and III were less widely distributed and contained mostly Australian isolates, whereas cluster IV consisted of only Australian isolates. Cluster V was unusual as it included one SP (SP25) with isolates from United States and China in the 1940s and 1950s and two SPs (SP26–SP28) containing mostly recent Japanese isolates. Cluster VI has only early isolates from 1920s to 1950s. There are considerable consistencies with MLVA data obtained previously (Kurniawan et al. 2010). Clusters I, II, III, IV, and V were represented by predominantly MT27, MT29, MT64, MT70, and MT186, respectively.

Thirteen SPs (SP1, SP6–SP9, SP11, SP18, SP20–SP22, SP32, SP40, and SP42) were not resolved into clusters. Of these, SP1, SP9, SP11, and SP18 contains isolates from two or more countries. These SPs were typically located at the base of a node without a branch length and potentially contained more diversity as indicated by the presence of multiple MTs (fig. 1). For SP1, SP9, and SP18, this was probably due to phylogenetic discovery bias (Pearson et al. 2004) as none was represented among the CGS isolates. SP11 was separated from SP42, represented by the CGS isolate PRCB588, by 1 of the 10 randomly selected terminal branch SNPs (supplementary fig. 1, Supplementary Material online), so we tested the remaining nine SNPs but found that none can give further differentiation of SP11.

**Evolution of Antigen Gene Variants of B. pertussis**

We typed the 287 isolates not typed previously (Kurniawan et al. 2010) for the genes *ptxA*, *fim3*, *fim2*, *fhaB*, and *prn* that code for ACV components. The 316 B. pertussis isolates were differentiated into 14 APs (fig. 2, table 2), 9 of which (AP1–AP4, AP7–AP9, AP12, and AP14) were shared by multiple isolates, whereas the remaining 5 (AP5, AP6, AP10, AP11, and AP13) each consisted of a single isolate. Three isolates, one each in AP1, AP3, and AP14, had a synonymous mutation in *fim3*, giving them a *fim3A* rather than *fim3A* allele. They were designated as AP1*, AP3*, and AP14*, respectively. Note that AP11 is not on the tree in Figure 2 as the isolate carrying AP11 cannot be fully typed by SNP typing.

The relationships between APs, based on eBURST analysis, are shown in Figure 2. The majority of APs were connected by a single allelic change, except for AP14 and the
AP12-AP13 pair, which differed by two or more alleles from the other APs and so are not connected to the other APs. AP9–AP14 were present only in the early isolates and presumably were the old more diverse allelic types. Seven other APs radiated from AP1 by single allelic changes. Five APs (AP2–AP6) differed from AP1 in the \textit{prn} allele, whereas AP7 differed from AP1 in the \textit{fim2} allele and AP9 in the \textit{ptxA} allele.

We mapped the APs to the SNP-based tree and found that these profiles correlated with evolutionary lineages (fig. 1). Clusters I and III contain mostly AP3 and AP1 with a small proportion of AP8 and AP6, respectively. Clusters II, IV, and V uniquely had AP4, AP7, and AP9, respectively. However, cluster VI contains multiple APs, suggesting that the old isolates were more antigenically diverse. All APs in clusters I–IV had the \textit{ptxA} allele, whereas cluster V had the \textit{ptxA} allele. This change from \textit{ptxA}2 to \textit{ptxA}1 seems to have been a major event because almost all recent isolates contain \textit{ptxA}1. Only a small proportion of recent isolates, all from cluster V, contained \textit{ptxA}2 as AP9. Based on the tree, AP9 is likely to be the ancestral AP giving rise to AP1, from which AP7 in cluster IV, AP4 in cluster II, and AP3 in cluster I arose with changes from \textit{fim2}-1 to \textit{fim2}-2, \textit{prn1} to \textit{prn3}, and \textit{prn1} to \textit{prn2}, respectively. There were several parallel changes observed: AP3 in SP9, SP1, and SP8 from \textit{prn1} to \textit{prn2}, AP4 in SP18 and SP21 from \textit{prn1} to \textit{prn3}, and AP9 in SP14 and SP16 from \textit{fim3A} to \textit{fim3B}. There is also a reverse change to AP1 in SP13 from \textit{prn2} to \textit{prn1}. AP1 is present in several unclustered SPs, but most cases are due to common ancestry rather than parallel changes because it is an ancestral AP.

**Discussion**

We previously analyzed sequence variation in the antigen genes of four major MTs in Australia and demonstrated...
changes consistent with selection pressure from vaccine immunity (Kurniawan et al. 2010). The SNP typing described in this paper has allowed the analysis of variation in antigen genes to be discussed in an evolutionary context. In Australia, WCV was used prior to 1997, followed by a transition period (1997–1999) when both WCV and ACV were used, and from 2000 onwards, only ACV has been used. A similar progression occurred in the other countries sampled. The period will vary and there may be some differences in the vaccines used, but the change from WCV to ACV occurred in all. There were a total of 78, 45, and 84 isolates from the WCV, transition, and ACV periods, respectively. The cluster data indicated that two clusters increased and two decreased in frequency from the WCV to ACV periods (fig. 3). Cluster I first emerged during the transition and increased to 31% in the ACV period, whereas cluster II decreased from 33.3% to 10.7%. Cluster III also decreased from 19.2% to 14.3%. Cluster IV increased markedly, from 5.1% in the WCV period to 32.1% in the ACV period. The frequency changes in clusters I, II, and IV are statistically significant.

The predominant ACV used in Australia (GlaxoSmithKline, Melbourne, Australia) contains PT, FHA, and PRN but no FIMs. The specific alleles in this ACV—Prn1, PtxA2, and FhaB1—originated from Tohama I. The second ACV, which is less frequently used in Australia, contains PT, PRN, and FHA, and also FIM2 and FIM3 (Sanofi Adacel). The likely Fim alleles in this ACV are Fim2-1 and Fim3A (Kurniawan et al. 2010). Thus, we have designated prn1, ptxA1, fhaB1, fim2-1, and fim3A as “ACV alleles” and others as “non-ACV alleles.” All isolates in clusters I–IV carry ptxA1 and fhaB1. Cluster I contained AP3 and AP8, which were associated with allelic change from prn1 to prn2 (AP3) and an additional change from fim3A to fim3B (AP8). Both changes are from ACV alleles to non-ACV alleles. Similarly, cluster IV carried AP7 with a non-ACV allele fim2-2. Cluster II was mainly associated with AP4, which carries prn3, a non-ACV allele that is apparently not associated with any selective advantage. Cluster III carries AP1 and presumably the ACV would select against it as all AP1 alleles except ptxA1 are ACV alleles. Therefore, we suggest that clusters I and IV have emerged as a result of ACV selection pressure.

The selective advantage of prn2 in cluster I and lack of advantage of prn3 in cluster II are supported by previous studies (Mooi et al. 1999; He et al. 2003). Prn3 differs from Prn1 by one repeat type with two amino acid differences in region 1 of the Prn molecule, whereas Prn2 differs from Prn1 by one repeat type and one extra repeat in region 1 (Mooi et al. 1999). Immunological responses to Prn are apparently type-specific (He et al. 2003). Infection by prn3-carrying isolates induced antibodies against the region 1-dependent epitopes of Prn1, but infection by prn2 isolates did not (He et al. 2003). The change in prn3 may not offer advantage against ACV but may offer other advantages such as better host cell attachment (Mooi 2010). A recent study by Komatsu et al. (2010) shows that the ACV is less protective against an isogenic Tohama I mutant carrying non-ACV prn2 and ptxA1 alleles in a mouse model as the mutant has a better survival rate in comparison to wild type Tohama I carrying ptxA2 and prn1 alleles.

Another key change within cluster I was in fim3, which evolved from fim3A into fim3B independently in SP14 and SP16. This important allelic difference may affect cross-protection as the nonsynonymous difference is located in a known surface epitope (Tsang et al. 2004). Fim3B may not be covered by ACV, although the effect is likely to be minor in Australia because ACVs containing Fim3 represent a small proportion of ACVs used. The main change in cluster IV is from Fim2-1 to Fim2-2, which may result from selection pressure.
ently coincided with the change from lower at 19.5%. The reduction in genetic diversity apparent around the world with little geographical restriction. Our survey of genome-wide SNPs in the 10 representative isolates from the same SPs in several countries suggests that \textit{B. pertussis} clones are likely to be circulating around the world with little geographical restriction.

Cluster I is not only a predominant clone in Australia but is also represented worldwide as it includes the most frequent PFGE types in Canada, Finland, France, and the United States. This has important implications for vaccination strategies applicable not only to Australia but also to countries using a similar ACV formulation. Our findings suggest that variants such as Prn2, Fim3B, but also to countries using a similar ACV formulation.

The wide geographical and temporal spread of isolates tested in the study enabled us to track the temporospatial distribution of \textit{B. pertussis} clones. Twelve SPs were present in two or more countries. Sampling bias was the likely explanation for many SPs being only observed in Australia but not elsewhere in the world because sample size from Australia was much larger. On the other hand, finding representative isolates from the same SPs in several countries suggests that \textit{B. pertussis} clones are likely to be circulating around the world with little geographical restriction.

Cluster I is not only a predominant clone in Australia but is also represented worldwide as it includes the most frequent PFGE types in Canada, Finland, France, and the United States. This has important implications for vaccination strategies applicable not only to Australia but also to countries using a similar ACV formulation. Our findings suggest that variants such as Prn2, Fim3B, and Fim2-2 should be considered for inclusion in the ACV.

Based on the CGS data from the 1920s UK isolate (NCTC10901) and the 1950s Japanese isolate (Tohama I), the SNP diversity of prevaccination and early WCV isolates was 37.5%. In comparison, the diversity of the recent isolates from multiple countries (Australia, Finland, France, and Japan), following the introduction of WCV, was much lower at 19.5%. The reduction in genetic diversity apparently coincided with the change from \textit{ptxA2} to \textit{ptxA1}. All recent isolates, except those in cluster V, carried \textit{ptxA1} (fig. 1). WCV was commonly produced from \textit{B. pertussis} Tohama I, which carries \textit{ptxA2}. Its continuous use may have driven the change to and encouraged the expansion of the \textit{ptxA1} lineage (clusters I–IV). Similarly, cluster I may have further expanded in response to ACV selection pressure, leading to further reduction in \textit{B. pertussis} population diversity.

Our survey of genome-wide SNPs in the 10 \textit{B. pertussis} genomes has confirmed previous observations that the level of SNP variation in \textit{B. pertussis} is low (Maharjan et al. 2008). The average frequency of polymorphism in \textit{B. pertussis} is one SNP for every 32 kb, which is significantly lower than reported for three other human clonal bacterial pathogens: \textit{Mycobacterium tuberculosis} with 1 per 3 kb (Fleischmann et al. 2002; Gutacker et al. 2002); \textit{Escherichia coli} O157:H7, with 1 per 6.7 kb (Zhang et al. 2006); and \textit{Salmonella enterica} serovar Typhi with 1 per 10 kb (Deng et al. 2003). On the other hand, the SNP frequency in \textit{B. pertussis} is similar to that in \textit{Mycobacterium leprae} (1 per 28 kb) (Monot et al. 2005) and marginally higher than that in \textit{Yersina pestis} (1 per 35 kb) (Chen et al. 2006). These data support the hypothesis that \textit{B. pertussis} was only recently derived (Diavatopoulos et al. 2005).

The high proportion of nsSNPs in the backbone genes does not mean that they are due to selection in this case as mildly deleterious mutations will not have been eliminated in the relatively short period since divergence, a phenomenon that has been observed in other groups of closely related strains (Rocha et al. 2006). Furthermore, the presence of many pseudogenes in its genome indicates that \textit{B. pertussis} has been undergoing reductive evolution (Parkhill et al. 2003) and deleterious mutations in some genes may not have been selected out if the gene product is no longer important. Clearly, some of the mutations could affect protein function as 26 of the 68 nsSNPs in the backbone genes were nonconservative changes based on analysis using the SIFT algorithm (Ng and Henikoff 2003), but we have no reason to implicate any specific mutations in selection. In contrast, the virulence genes are likely to be under selection pressure for change. All but 3 of the 16 SNPs in virulence-associated genes tested, the majority of which encode significant antigens, are nsSNPs (supplementary table S2, Supplementary Material online), which is higher (81%) than that in backbone genes (61%), although only three are nonconservative changes. Six of the 13 nsSNPs have been reported in L517 previously, including three (one each in \textit{ptxA}, \textit{prn}, and \textit{tcfA}) in both L517 and other strains (van Loo et al. 2002; Packard et al. 2004), and the remaining

### Table 2. Characteristics of the APs.

<table>
<thead>
<tr>
<th>AP Number</th>
<th>ptxA</th>
<th>fim3</th>
<th>fim2</th>
<th>fhaB</th>
<th>prn</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP1</td>
<td>1</td>
<td>A</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AP1*</td>
<td>1</td>
<td>A*</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AP2</td>
<td>1</td>
<td>A</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>AP3</td>
<td>1</td>
<td>A</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>AP3*</td>
<td>1</td>
<td>A*</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>AP4</td>
<td>1</td>
<td>A</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>AP5</td>
<td>1</td>
<td>A</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>AP6</td>
<td>1</td>
<td>A</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>AP7</td>
<td>1</td>
<td>A</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AP8</td>
<td>1</td>
<td>B</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>AP9</td>
<td>2</td>
<td>A</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AP10</td>
<td>2</td>
<td>A</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>AP11</td>
<td>2</td>
<td>A</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>AP12</td>
<td>4</td>
<td>A</td>
<td>2</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>AP13</td>
<td>4</td>
<td>A</td>
<td>2</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>AP14</td>
<td>5</td>
<td>A</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>AP14*</td>
<td>5</td>
<td>A*</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

Note.—The APs marked with asterisks are profiles with a synonymous change in \textit{fim3A}.

![Figure 3](image-url)
three in L517 only (one each in bvgs, fimD, and ptxB) (Maharjan et al. 2008). Fimbrial genes seem to be particularly variable, with variation in three genes. Fimbriae are involved in tracheal colonization (Geuijen et al. 1997) and are composed of a major subunit, Fim2 and/or Fim3, and FimD a minor subunit located at the tip of the fimbria. The major subunit determines serotype specificity. Variation in fim2 and fim3 has been reported in many studies (Packard et al. 2004; Tsang et al. 2004). fimD was found to vary in L517 by CGS analysis previously (Maharjan et al. 2008) and is now seen in the other 10 CGS isolates, including the most divergent isolate, 18323. Two new nsSNPs were found in this study, one each in fim2 (L500) and fim3 (FR0743 and L900). SNPs have also been observed in genes that code for proteins homologous to FHA, fhaS (FHA-like, small), and fhaL. FHA-like, large). Their functions in colonization have not been determined, but these genes have been shown to be well expressed, although at much lower levels than fhaB (Antoine et al. 2000).

Conclusions

In this study, we identified 45–70 SNPs each in a total of 1.4 Mb of the 4.1 Mb genome interrogated using CGS from 10 spatiotemporally diverse B. pertussis strains, confirming the limited genetic diversity of B. pertussis. The genome-wide SNPs were used to classify a collection of 316 isolates from five countries. They were differentiated into 42 SPs, most of which grouped into six clusters. The emergence and increase in prevalence of clusters I and IV in Australia, with different APs, provided new and important evidence of selection pressure from vaccination on the evolution of B. pertussis. Cluster I has been identified as a major clone with a worldwide distribution. Most strains of B. pertussis currently causing clinical pertussis have apparently descended from a single prevaccine lineage. Expansion of clones carrying non-ACV alleles is temporally related to, and possibly driven by, the introduction of ACVs. These findings have significant implications for control of pertussis and vaccination strategies.

Supplementary Material

Supplementary tables S1–S3 and supplementary figs. S1–S3 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We are indebted to the following researchers and clinicians for the generous donations of isolates: Dr Kazunari Kama- chi, Department of Bacterial Pathogenesis and Infection Control, National Institute of Infectious Diseases, Tokyo, Japan; Dr Shane Byrne, Sullivan Nicolaides Pathology, Queensland, Australia; Dr John Tapsall, Prince of Wales Hospital, Australia; Dr Ian Carter, St George Hospital, Australia; Dr Margret Ip, Chinese University of Hong Kong; Dr Qishui He, Pertussis Reference Laboratory, National Public Health Institute, Turku, Finland; Dr Nicole Guiso, Institut Pasteur, France; Dr Raymond Tsang, National Micro-

References


