Significant Gene Order and Expression Differences in *Bordetella pertussis* Despite Limited Gene Content Variation†

Mary M. Brinig,[1,6]* Craig A. Cummings,[1,6] Gary N. Sanden,[3] Paola Stefanelli,[4] Andrew Lawrence,[5] and David A. Relman[1,2,6]

Departments of Microbiology and Immunology[1] and Medicine,[2] Stanford University School of Medicine, Stanford, California 94305; Centers for Disease Control and Prevention, Atlanta, Georgia 30333[3]; Department of Infectious, Parasitic and Immune-Mediated Diseases, Istituto Superiore di Sanità, 00161 Rome, Italy[4]; Women's and Children's Hospital, North Adelaide, South Australia 5006, Australia[5]; and VA Palo Alto Health Care System, Palo Alto, California 94304[6]

Bordetella pertussis, an obligate human pathogen and the agent of whooping cough, is a clonal species, despite the dynamic selection pressures imposed by host immunity and vaccine usage. Because the generation of variation is critical for species evolution, we employed a variety of approaches to examine features of *B. pertussis* genetic variation. We found a high level of conservation of gene content among 137 *B. pertussis* strains with different geographical, temporal, and epidemiological associations, using comparative genomic hybridization. The limited number of regions of difference were frequently located adjacent to copies of the insertion element IS*481*, which is present in high numbers in the *B. pertussis* chromosome. This repeated sequence appears to provide targets for homologous recombination, resulting in deletion of intervening sequences. Using subtractive hybridization, we searched for previously undetected genes in diverse clinical isolates but did not detect any new genes, indicating that gene acquisition is rare in *B. pertussis*. In contrast, we found evidence of altered gene order in the several strains that were examined and again found an association of IS*481* with sites of rearrangement. Finally, we compared whole-genome expression profiles of different strains and found significant changes in transcript abundance, even in the same strain after as few as 12 laboratory passages. This combination of approaches provides a detailed picture of a pathogenic species with little gene loss or gain but with the capacity to generate variation by rearranging its chromosome and altering gene expression. These findings have broad implications for host adaptation by microbial pathogens.

Bacterial pathogens face a dynamic array of selective pressures as they establish themselves within hosts, replicate, and are transmitted to new hosts. Upon arrival in a host, microbes may be forced to contend with innate immune defenses, scarcity of metabolic resources, competition from other bacteria for resources and colonization sites, and adaptive immunity, and then they must readapt to the outside environment during transmission to a new host. Individual hosts and pathogens adapt and respond to each other during the course of an infection, while evolving new offensive and defensive strategies on a population-wide basis. Pathogens must employ strategies to prevent their eradication by host immunity, which is constantly changing both within an individual host and within the host population. For an obligate pathogen to persist in a host population, its evolutionary process must be ongoing. This evolution can include gaining new functionality through horizontal transfer of genes, losing "antivirulence" genes, or altering the form or regulation of existing proteins through mutation in coding sequences or changes in promoters or operonic structures.

The *Bordetella* genus of respiratory pathogens provides a rich model with which to explore questions of evolution of pathogenesis and host adaptation. The three species comprising the "classical" bordetellae have much in common, including very similar mechanisms of pathogenesis and a high degree of sequence similarity in shared genes, but they have different host ranges and abilities to survive outside a host. *Bordetella bronchiseptica*, which has a broad mammalian host range and the most complete set of metabolic pathways of the three species, is thought to be similar to the ancestral form of the group from which the more-host-restricted variants evolved by genome reduction (7, 28). *Bordetella parapertussis* infects only sheep and humans, and *Bordetella pertussis* is an obligate pathogen of humans.

B. pertussis causes whooping cough, a significant source of mortality against which vaccines have been widely used in developed countries for approximately 50 years. The medical importance of this pathogen has motivated numerous molecular and epidemiological investigations which have revealed several distinctive features of this species. One hallmark of *B. pertussis* is its extremely limited genetic variability, as assessed by multilocus enzyme electrophoresis (26, 37), multilocus sequence typing (8, 38), and previous microarray-based comparative genomic hybridization (CGH) analysis performed on a small number of strains (7). Because *B. pertussis* induces a long-lasting immune response in its hosts, which theoretically imposes selective pressure on the bacterium to evade immunity induced by prior infection and vaccination, the lack of polymorphism in the species is somewhat surprising and has been speculatively attributed to fitness costs associated with immune evasion or to frequent population bottlenecks (36).

Another unusual feature of *B. pertussis* is its exceptionally high load of insertion sequence (IS) elements. The sequenced strain, Tohama I, carries 261 IS elements, 238 of which are identical repeats of IS481 (28). These elements comprise over 6% of the genome and have transposed into many coding sequences, creating numerous presumably nonfunctional pseudogenes in Tohama I and also providing sites for homologous recombination throughout the chromosome. The IS elements appear to have expanded recently in the *B. pertussis* genome, as many fewer IS elements are found in *B. bronchiseptica* and *B. parapertussis* strains (28, 37).

As well as the loss of functional genes due to disruption by IS elements, several lines of evidence have led to the inference that *B. pertussis* has undergone significant genome reduction (7, 28). In addition, very little evidence of horizontal gene acquisition was observed in the genome sequence of the single strain so far examined (7, 28). To gain a better understanding of how this pathogen is evolving to meet the pressures of its restricted host niche without the benefit of novel genetic input, we obtained strains that represented a wide diversity of temporal and geographic isolation characteristics, as well as strains from before and during a large epidemic, and strains from hosts with different vaccination status. We subjected these strains to CGH, subtractive hybridization, global expression profiling, and gene order analysis. Consistent with previous studies, we found very little variation in gene content between strains. However, variation in gene order may provide a source of genetic variability necessary for evolution of this species. Additionally, we found that whole-genome expression profiles of a recent clinical strain changed significantly over 12 laboratory passages, indicating that despite extremely restricted variability in gene content, *B. pertussis* can alter gene regulation quickly when introduced to a new environment.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *Bordetella* strains used in this study are listed in Table S1 of the supplemental material. Strains individually discussed in this report are as follows: Tohama I, isolated in Japan in the 1950s and sequenced at the Sanger Centre (28); 18323, isolated in the United States in the 1940s and used as the ATCC type strain; Bpe43, Bpe336, and Bpe337, all isolated during an epidemic in Cincinnati, Ohio, in 1993; Bpe280, isolated in California in 2005. Liquid cultures were grown in modified Stainer-Scholte medium at 37°C with shaking. For passage experiments, strains were grown at 37°C on Bordet Gengou agar with 15% sheep blood (BD Biosciences, San Jose, CA) until individual colonies were well-defined, and a single colony was streaked onto a fresh plate.

Microarray design and construction. An expanded *Bordetella* microarray was constructed based on that described in reference 7. A total of 5,670 PCR products, representing 97.4% of the *B. pertussis* Tohama I open reading frames (ORFs), 97.9% of the *B. parapertussis* 12822 ORFs, and 98.5% of the *B. bronchiseptica* RB50 ORFs, were printed in duplicate on poly-L-lysine-coated glass slides. Details of microarray printing have been previously reported (7).

Comparative genomic hybridization. Genomic DNA from each strain was labeled and hybridized to the arrays along with mixed reference DNA (consisting of equimolar quantities of genomic DNA from each of the three sequenced strains used in construction of the array). Technical details are in the methods of the supplemental material. The threshold for calling an array element "not detected" was determined individually for each array based on its distribution of log ratios, as previously described (7). Regions of difference (RDs) were defined as two or more adjacent array elements (separated by a maximum gap of one probe) not detected in at least 3 (of 137) strains. Data have been deposited in ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) with accession number E-TABM-54.

Subtractive hybridization. Suppression subtractive hybridization was carried out using the PCR-Select bacterial genome subtraction kit (Clontech, Mt. View, CA), with modifications as previously described (7). The tester DNA was a pool

of equal amounts of genomic DNA from five *B. pertussis* strains chosen to represent different geographical locations and time periods (Bpe37, Bpe43, Bpe183, WCH22, and Bp12464). A control tester pool was prepared by adding an equimolar amount of HaeIII-digested ϕX174 phage DNA after completing the initial RsaI digestion step of the protocol. The driver DNA was a pool of equimolar amounts of genomic DNA from the sequenced strains *B. pertussis* Tohama I, *B. parapertussis* 12822, and *B. bronchiseptica* RB50, as well as 40 fragments isolated from an ovine isolate of *B. parapertussis* that were not found in the sequenced *Bordetella* strains (data not shown). From the subtraction using the control tester pool, 52 randomly chosen fragments were sequenced. From the subtraction using the experimental tester pool, 480 fragments were recovered, cloned, reamplified, and printed on microarrays. The microarrays were hybridized with each of the five strains in the tester pool versus the driver pool, and the 10 spots from each array with the highest ratios of tester versus driver were chosen for sequencing.

Microarray-based chromosomal mapping. Genomic DNA was prepared in agarose plugs, digested with SwaI, and resolved by pulsed-field gel electrophoresis (PFGE) essentially as described elsewhere (25; see also the Methods section in the supplemental material). Each band was excised and treated with β-agarase (New England Biolabs, Ipswich, MA) according to the manufacturer's instructions. DNA from each band was isopropanol precipitated in the presence of 30 μM spermine, 70 μM spermidine, and 100 mM NaCl and then labeled and hybridized to microarrays as described in the methods section of the supplemental material.

Analysis of transcript abundance profiles. Mid-log-phase cultures were diluted in fresh Stainer-Scholte medium to a calculated absorbance of 0.05 at 600 nm. Half of each culture was removed to another flask, and MgSO$_4$ was added to a final concentration of 75 mM to modulate the strains into the Bvg$^-$ phase. When cultures reached mid-log phase (optical density at 600 nm, 0.8 to 2.6), 1 ml of culture was centrifuged at 16,000 × *g* for 30 seconds, the supernatant was removed, and the cell pellet was frozen at −80°C. Total RNA was isolated using the RNAqueous-4PCR kit (Ambion, Austin, TX) with an additional lysis step using lysozyme (0.4 mg/ml). RNA labeling and hybridization are described in the methods section of the supplemental material. Significance analysis of microarrays (SAM) (35) was performed using all spots with data in at least 67% of strains (triplicate cultures were averaged together).

RESULTS

Analysis of gene content of diverse *B. pertussis* strains by comparative genomic hybridization. To gain a broad perspective on the gene content of *B. pertussis*, a DNA microarray was used to survey genomic DNA from 137 *B. pertussis* strains isolated from four countries, including strains from before, during, and after an epidemic in Cincinnati, Ohio, and from before and after high vaccination coverage was achieved in the United States and The Netherlands. Comparisons of the gene content of these *B. pertussis* strains revealed very low levels of intraspecies diversity. On average, only 1.6% (standard deviation, 0.62%) of the total pool of genes found in any *B. pertussis* strain were called "not detected." Figure 1 shows the comparative genomic hybridization data for all 137 strains. The materials section in the supplemental material includes two microarray data files, one containing the log ratios for all array elements found in *B. pertussis* strains and one with binary data consisting of "detected" or "not detected" calls for each array element, with RDs indicated.

Twenty-six loci fit our criteria for RDs (at least two chromosomally adjacent array elements not detected in three or more strains). The genes in 20 RDs were detected in Tohama I; the other 6 RDs contained *Bordetella* genes that were absent in Tohama I but present in other *B. pertussis* strains. Four of the RDs which contained genes detected in Tohama I were composed of array elements with slightly lower ratios in many strains, which appear to be divergent regions as opposed to frank deletions. These four RDs are not flanked by IS elements, whereas all
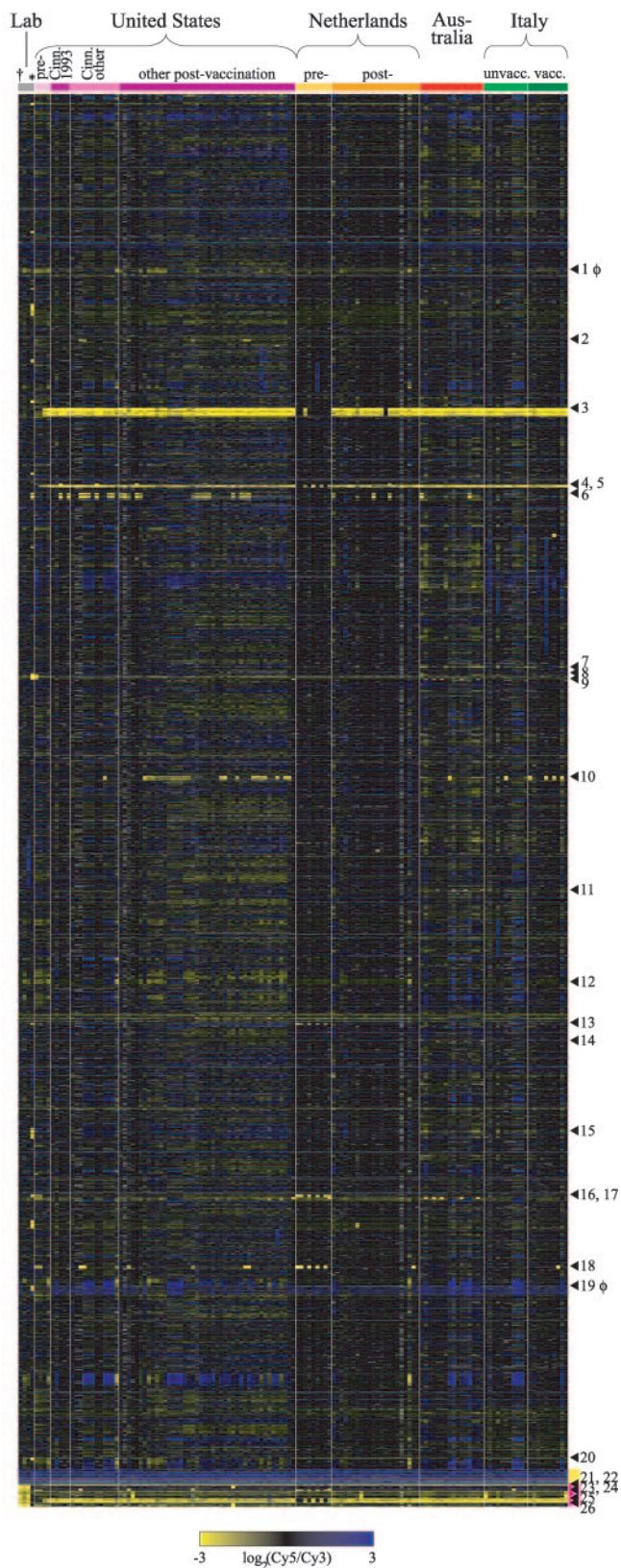
FIG. 1. CGH data for 137 *B. pertussis* strains, arranged by collection. Abbreviations: lab, laboratory strains; pre-, prevaccine era (1935 to 1958); post-, postvaccine era (1977 to 2005); Cinn. 1993, from a 1993 outbreak in Cinncinati, Ohio; Cinn. other, from Cinncinati between

TABLE 1. Functional categories of genes under- or overrepresented in regions of difference[d]

Category[a]	No. of array elements (% of total)	
	Tohama I genome	RDs
Pseudogenes	538 (14.7)	45 (23.2)
Adaptation (includes iron storage and osmotic adaptation)	67 (1.8)	9 (4.6)
Laterally acquired elements	63 (1.7)	52 (26.8)
Metabolism of small molecules[b]	613 (16.8)	15 (7.7)
Macromolecule metabolism, ribosome[c]	304 (8.3)	5 (2.6)

[a] All categories except pseudogenes are single or combined FC headings (reference 31, as classified in reference 28).
[b] Includes the headings amino acid biosynthesis; biosynthesis of cofactors, carriers; central intermediary metabolism; degradation of small molecules; energy metabolism, carbon; fatty acid biosynthesis; and nucleotide biosynthesis.
[c] Includes the headings macromolecule degradation; macromolecule synthesis, modification; and ribosome constituents.
[d] $P < 0.05$, binomial distribution with Bonferroni correction.

of the other RDs either are adjacent to at least one IS*481* element in the Tohama I sequence or there is an IS*481* element in place of the RD if it is missing from Tohama I, suggesting that a major mechanism for gene loss in *B. pertussis* is IS element-mediated deletion. There also were clear instances of deletions that did not meet the criteria for RDs because they involved a single gene missing from multiple strains or several genes missing from one or two strains, and almost all of these regions were adjacent to one or two IS*481* elements.

Pseudogenes were overrepresented in those RDs that consisted of Tohama I genes (Table 1). In addition, two RDs (RD1 and RD19) appeared to be decaying prophages. Genes in the functional category "adaptation," which includes those involved in iron uptake and copper resistance, also were overrepresented in the RDs, suggesting that these genes may be unused or possibly even detrimental to these bacteria in some situations (see Table 1 for explanation of functional categorization of genes). The categories "metabolism of small molecules" and "macromolecule metabolism and ribosomal genes" were significantly underrepresented in the RDs, indicating that these genes are under selection to be retained, as expected given their essential functions. Interestingly, although the chemotaxis and motility loci in Tohama I contain several pseudogenes and IS*481* elements and *B. pertussis* motility has never been observed (3), no genes in this functional category appeared in any of the RDs, although the number of genes was too small for this underrepresentation to reach statistical signifi-

1989 and 1994 but not during the 1993 outbreak; unvacc., unvaccinated patients; vacc., vaccinated patients. Each column represents a strain; each row represents a microarray element (usually the average of duplicates), arranged in Tohama I genome order from the top of the figure to the yellow bar. The dagger designates Tohama I; the asterisk designates strain 18323 (both in the lab strains collection). Arrowheads indicate RDs, defined as two or more adjacent genes not detected in three or more strains. RDs consisting of prophage genes are marked with ϕ next to the number. Yellow bar, genes duplicated in Tohama I. Magenta bar, genes not present in the Tohama I sequence, arranged in *B. bronchiseptica* RB50 order. Missing data are gray (display excludes genes with missing data from more than 25% of arrays).

cance. Notably, no known virulence factors were found among the RDs.

Three RDs were associated with strains isolated in the prevaccine era from the United States and The Netherlands. The genes in RD3, comprising BP0911 to BP0934, were present in 11 of 14 strains from the prevaccine era but only 1 of 82 strains from the postvaccine era in these two countries ($P < 1 \times 10^{-10}$; Fisher's exact test, two-tailed). Among the 24 genes in this RD, one encodes a putative membrane protein and two encode putative exported proteins, which are potential targets of vaccine-induced immunity. A similar temporal distribution was observed for RD5, encompassing BP1135 to BP1141, which includes four genes encoding proteins predicted to be involved in iron uptake. These genes were detected in 10 of 16 prevaccine strains from the United States, The Netherlands, and Japan, whereas these genes were not detected in any postvaccine strains from these countries or in modern strains from vaccinated and unvaccinated Italian patients ($n = 121$; $P < 1 \times 10^{-10}$). Finally, RD25, which contains genes involved in fatty acid metabolism (BB1789 to BB1797), was detected in 8 of 16 prevaccine strains and 0 of 121 postvaccine strains ($P < 1 \times 10^{-8}$).

The opposite trend was observed for RD13 (BP2627-2629), RD16 (BP3104-3109), RD18 (BP3314-3322), and RD23 (BB0917-0921), all of which were detected predominantly in strains collected in the postvaccine era. In each case, some prevaccine strains did contain the genes in these RDs, suggesting that the deleted strains were replaced by other strains circulating at the time that contained these genes.

The only strain with unusually high divergent gene content was the ATCC type strain 18323, which also has been differentiated from most *B. pertussis* strains by other typing methods (5, 23, 26, 40). This strain had the largest proportion of genes called "not detected" (5.3%), and the locations of not-detected regions frequently did not align with such regions in any other strain. In addition, strain 18323 contains genes found in other *Bordetella* species but not present in any other *B. pertussis* strains (5, 23, 26, 40). The fact that this strain is missing genes that are present in every other *B. pertussis* strain surveyed indicates that such genes are not essential for survival of the bacterium but may be important for achieving high levels of circulation within human populations (only one other *B. pertussis* strain similar to 18323 has been reported [5, 40]).

Investigation of *B. pertussis* gene acquisition by subtractive hybridization. A limitation of microarray-based CGH is that it can detect only genes that are present in the strain(s) used to construct the microarray. Therefore, to determine if other strains contain additional DNA, we performed subtractive hybridization. Genomic DNA from five *B. pertussis* strains from different countries and time periods was pooled and subjected to suppression subtractive hybridization, using a driver pool of DNA from the sequenced strains of *B. pertussis*, *B. parapertussis*, and *B. bronchiseptica*, supplemented with sequences from an ovine isolate of *B. parapertussis*. Four hundred eighty recovered fragments were screened by microarray hybridization analysis to find those most likely to be unique to the tester pool and 36 of these were sequenced, but BLASTN analysis indicated that all of them shared very high (or complete) nucleotide identity with sequences found in the driver pool. Thus, no novel sequences were recovered from the tester pool.

To gain an estimate of the sensitivity of this technique, the tester DNA pool was spiked with an equimolar quantity of HaeIII-digested φX174 phage DNA. Seventy-two percent of the φX174 spiked DNA (3,858/5,386 bp; 6 of 11 fragments) was recovered after sequencing 52 randomly chosen fragments from this subtraction. Because the experimental subtractive hybridizations included an additional screening step, we expect that the technique should be sensitive enough to detect most unique *B. pertussis* genes, and almost certainly any large insertions, that are present in the tester sequences.

Detection of genome rearrangements by microarray-based chromosomal mapping. Recombination between the many IS elements in *B. pertussis* may result in chromosomal deletions and inversions, depending on the orientation and location of the repeated sequences. Such chromosomal rearrangements are common in *B. pertussis* (32, 33). Many of the deletions observed by CGH appear to be the product of recombination between two IS elements that were in the same orientation on the chromosome. We employed a novel technique to investigate the extent of genome rearrangement and the role of IS elements in mediating it. Genomic DNA was digested into large fragments, and the gene content of each fragment was determined using microarray hybridization. SwaI digestion of Tohama I, the sequenced strain of *B. pertussis*, produced five bands of the predicted sizes, as assessed by PFGE. When DNA from each of these bands was hybridized to microarrays, the array elements detected showed excellent correspondence with the genes expected to be on each fragment according to the genome sequence, confirming the accuracy of this technique (data not shown).

Digestion with SwaI of genomic DNA from independently isolated *B. pertussis* strains resulted in a remarkable diversity of PFGE banding patterns (Fig. 2, top panel). Determination of the genes present in each of the bands produced by SwaI digestion of three isolates, Bpe43, Bpe336, and Bpe337, revealed that they are distinguished from Tohama I by several genomic rearrangements (Fig. 2, bottom panel). Each of the bands from these strains contained one to four blocks of genes which were contiguous in the Tohama I genome, indicating that multiple chromosomal inversions had occurred during the divergence of these strains to produce different arrangements of the same genetic material. At each apparent recombination junction, an IS481 element was present in the sequence of Tohama I, providing strong evidence that most of the chromosomal rearrangements are mediated by IS elements.

To investigate the rate of chromosomal rearrangements, we passaged two strains on plates and compared the gene order of the ancestral and derived isolates (Fig. 2, bottom panel). After 224 passages of Tohama I, a laboratory-adapted strain for which the precise passage history is not known, the gene order was unchanged from that of the original sequenced strain at the resolution provided by this technique. Laboratory adaptation or the absence of natural selection pressures may have affected the extent or frequency of genomic rearrangement in this strain. To determine whether a more recent clinical isolate might undergo rearrangement more frequently, we analyzed the gene order of Bpe280, which had been minimally passaged (an estimated five passages) before we obtained it. The gene order of Bpe280 was different from that of the other *B. pertussis* strains analyzed, but no rearrangements between the ancestral strain and a derivative which had been passaged 12
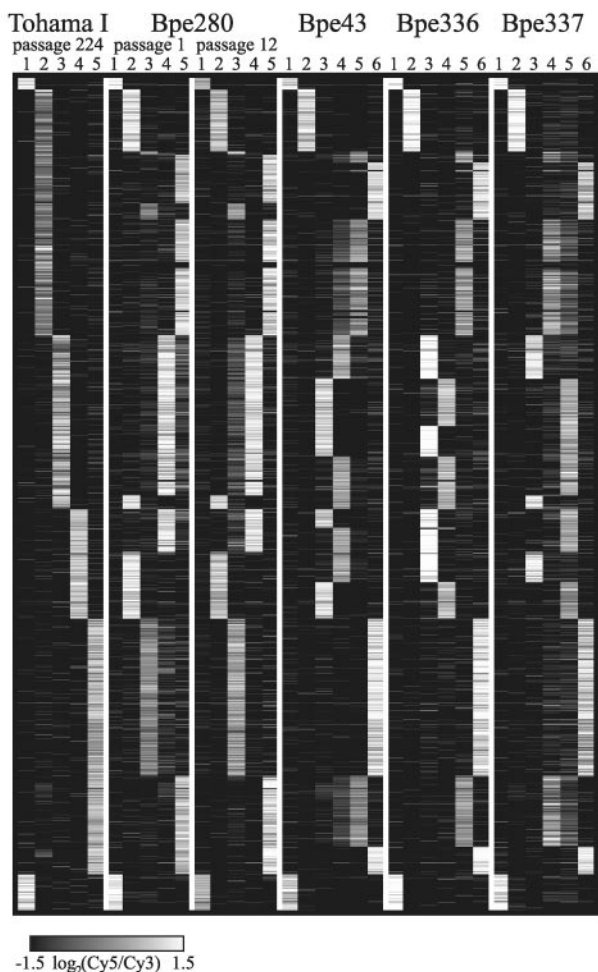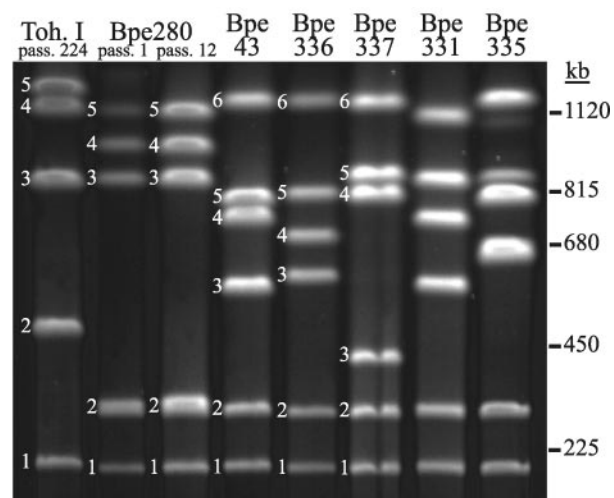
FIG. 2. Top: PFGE of SwaI-digested genomic DNA of seven strains, listed across the top of the gel. Numbered bands correspond to columns in the bottom panel. Bottom: Array elements detected on each SwaI fragment of six strains. Each column represents one SwaI fragment, and each row represents a microarray element (only those present in Tohama I and with data from at least 90% of all arrays), arranged in Tohama I passage 1 gene order (identical to Tohama I passage 224 order). Array elements present in a fragment are white, those absent are black, and missing data are gray.

times were detected (Fig. 2, bottom panel). Rearrangements that were within the span of a single SwaI fragment would have been undetectable by this method.

Analysis of differences in transcript abundance and Bvg regulation between pre- and postvaccine strains and laboratory-passaged strains. CGH and subtractive hybridization indicated that gene content was very similar among all isolates, but most intergenic regions are not represented on the array and neither method is able to detect small genetic alterations (e.g., point mutations, frameshifts, and short insertions or deletions) which could produce significant changes in expression levels. Indeed, several studies have found variation in gene expression even among strains with similar gene content (11, 12, 16, 24, 39). In addition, IS-mediated chromosomal rearrangements could cause genes to be moved to positions near different transcriptional regulatory motifs, resulting in different expression patterns.

To look for correlations between expression profiles and epidemiology of *B. pertussis*, we chose four Dutch strains from the prevaccine era and four from the postvaccine era, since these collections showed the most variation in gene content (although still quite limited). Only nine array elements were found to differentiate between the two groups of strains (determined by SAM; false discovery rate of 0% [data not shown]). These genes did not group together into apparent operons or categories of functional significance, and there were many more strain-specific differences than differences between the prevaccine and postvaccine groups. Therefore, we next chose to focus on a smaller number of strains with known relationships to each other.

The ancestral and passaged descendant (224 passages on plates) of Tohama I, a laboratory strain, showed very few differences in transcript abundance under Bvg$^+$ (virulent) and Bvg$^-$ (avirulent) conditions (four genes with a 20% false discovery rate and 10 genes with a 9% false discovery rate, respectively). In contrast, there were 38 and 41 genes (false discovery rate < 2%) with different transcript abundances in the ancestral and passaged strains of the clinical isolate Bpe280 under Bvg$^+$ and Bvg$^-$ conditions, respectively, after only 12 passages (Fig. 3). Of particular interest, transcripts of two genes involved in capsule biosynthesis were detected at lower levels in the passaged strain under Bvg$^-$ conditions, which may provide an explanation for the conflicting reports about whether *B. pertussis* produces a capsule (see the discussion in reference 28). Similarly, transcripts of several genes for chemotaxis and motility were present at lower levels in the passaged strain under Bvg$^+$ conditions, although it is not known whether these proteins are expressed. The *fim3* gene, encoding an adhesin, also had lower transcript levels in the passaged strain of Bpe280, and the same pattern was seen in Tohama I when a lower stringency was used for determining expression differences.

Surprisingly, genes for two virulence factors, pertactin (*prn*) and filamentous hemagglutinin (*fhaB*), had higher transcript abundance in the passaged strain under Bvg$^+$ conditions, as did a putative type III protein (*bscD*) under Bvg$^-$ conditions.

DISCUSSION

Bacterial species exhibit a wealth of mechanisms for generating genetic diversity, which is essential for adaptation to environmental changes, including the many dynamic challenges that confront

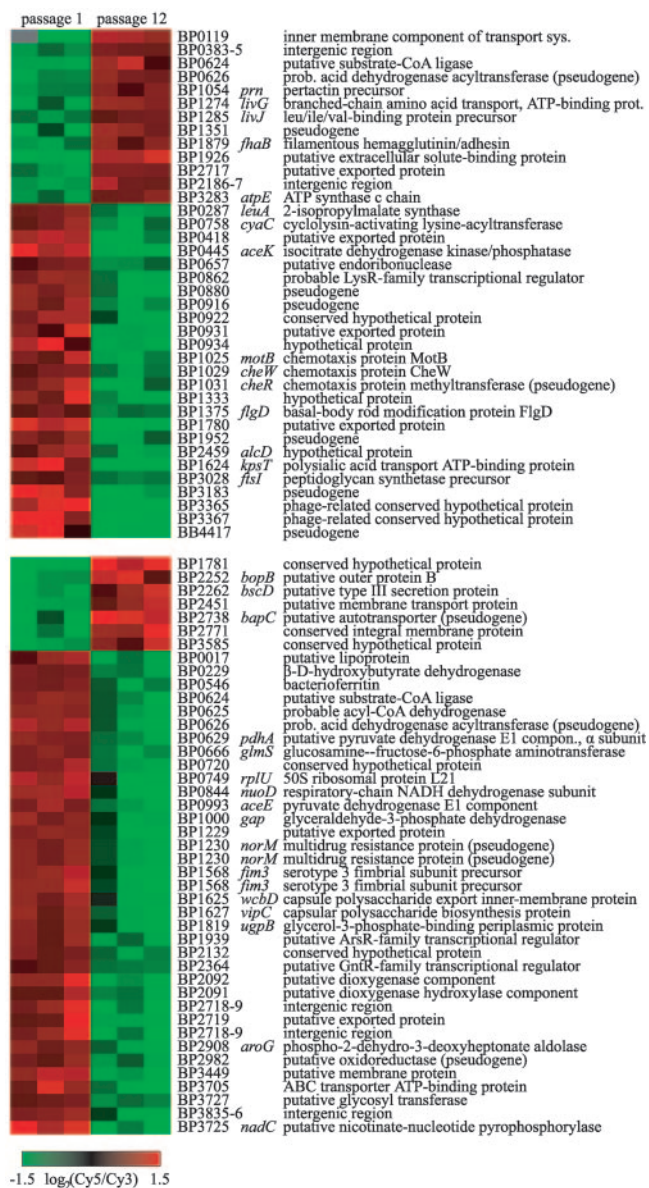| passage 1	passage 12		
BP0119		inner membrane component of transport sys.
BP0383-5		intergenic region
BP0624		putative substrate-CoA ligase
BP0626		prob. acid dehydrogenase acyltransferase (pseudogene)
BP1054	prn	pertactin precursor
BP1274	livG	branched-chain amino acid transport, ATP-binding prot.
BP1285	livJ	leu/ile/val-binding protein precursor
BP1351		pseudogene
BP1879	fhaB	filamentous hemagglutinin/adhesin
BP1926		putative extracellular solute-binding protein
BP2717		putative exported protein
BP2186-7		intergenic region
BP3283	atpE	ATP synthase c chain
BP0287	leuA	2-isopropylmalate synthase
BP0758	cyaC	cyclolysin-activating lysine-acyltransferase
BP0418		putative exported protein
BP0445	aceK	isocitrate dehydrogenase kinase/phosphatase
BP0657		putative endoribonuclease
BP0862		probable LysR-family transcriptional regulator
BP0880		pseudogene
BP0916		pseudogene
BP0922		conserved hypothetical protein
BP0931		putative exported protein
BP0934		hypothetical protein
BP1025	motB	chemotaxis protein MotB
BP1029	cheW	chemotaxis protein CheW
BP1031	cheR	chemotaxis protein methyltransferase (pseudogene)
BP1333		hypothetical protein
BP1375	flgD	basal-body rod modification protein FlgD
BP1780		putative exported protein
BP1952		pseudogene
BP2459	alcD	hypothetical protein
BP1624	kpsT	polysialic acid transport ATP-binding protein
BP3028	ftsI	peptidoglycan synthetase precursor
BP3183		pseudogene
BP3365		phage-related conserved hypothetical protein
BP3367		phage-related conserved hypothetical protein
BB4417		pseudogene
BP1781		conserved hypothetical protein
BP2252	bopB	putative outer protein B
BP2262	bscD	putative type III secretion protein
BP2451		putative membrane transport protein
BP2738	bapC	putative autotransporter (pseudogene)
BP2771		conserved integral membrane protein
BP3585		conserved hypothetical protein
BP0017		putative lipoprotein
BP0229		β-D-hydroxybutyrate dehydrogenase
BP0546		bacterioferritin
BP0624		putative substrate-CoA ligase
BP0625		probable acyl-CoA dehydrogenase
BP0626		prob. acid dehydrogenase acyltransferase (pseudogene)
BP0629	pdhA	putative pyruvate dehydrogenase E1 compon., α subunit
BP0666	glmS	glucosamine--fructose-6-phosphate aminotransferase
BP0720		conserved hypothetical protein
BP0749	rplU	50S ribosomal protein L21
BP0844	nuoD	respiratory-chain NADH dehydrogenase subunit
BP0993	aceE	pyruvate dehydrogenase E1 component
BP1000	gap	glyceraldehyde-3-phosphate dehydrogenase
BP1229		putative exported protein
BP1230	norM	multidrug resistance protein (pseudogene)
BP1230	norM	multidrug resistance protein (pseudogene)
BP1568	fim3	serotype 3 fimbrial subunit precursor
BP1568	fim3	serotype 3 fimbrial subunit precursor
BP1625	wcbD	capsule polysaccharide export inner-membrane protein
BP1627	vipC	capsular polysaccharide biosynthesis protein
BP1819	ugpB	glycerol-3-phosphate-binding periplasmic protein
BP1939		putative ArsR-family transcriptional regulator
BP2132		conserved hypothetical protein
BP2364		putative GntR-family transcriptional regulator
BP2092		putative dioxygenase component
BP2091		putative dioxygenase hydroxylase component
BP2718-9		intergenic region
BP2719		putative exported protein
BP2718-9		intergenic region
BP2908	aroG	phospho-2-dehydro-3-deoxyheptonate aldolase
BP2982		putative oxidoreductase (pseudogene)
BP3449		putative membrane protein
BP3705		ABC transporter ATP-binding protein
BP3727		putative glycosyl transferase
BP3835-6		intergenic region
BP3725	nadC	putative nicotinate-nucleotide pyrophosphorylase

-1.5 $\log_2$(Cy5/Cy3) 1.5

FIG. 3. Array elements with significantly different levels of transcript abundance (determined by SAM) between the ancestral and passaged strains of clinical isolate Bpe280 during growth under Bvg$^+$ (top) or Bvg$^-$ (bottom) conditions. Each column represents one of the triplicate cultures of each strain. In some cases, data for multiple different array elements corresponding to the same sequence are shown. Data are mean centered for each array element, and array elements are arranged by hierarchical clustering. Missing data are gray.

pathogens. Genetic variation in pathogenic species may arise through uptake of new genetic material, which can confer traits such as antibiotic resistance, toxin production, and adhesion to host cells, or through gene loss during adaptation to a particular niche or lifestyle (18, 22). CGH studies of some pathogenic bacterial populations have revealed extensive variation in gene content, with over one-fifth of the content of some genomes being classified as "accessory genes" (9, 10, 30). The theme of highly flexible genome contents, due to both horizontal transfer and genomic degradation, has been reiterated in other CGH analyses

and plays a major role in our understanding of how pathogenic bacteria adapt.

However, not all bacterial pathogens exhibit substantial variation in gene content. Some species, such as *Pseudomonas aeruginosa*, maintain large genomes which confer adaptability to diverse environments (41). Other species, such as *Yersinia pestis* and *B. pertussis*, have evolved from more-generalist ancestors primarily through extensive genome degradation (1, 28). While gene acquisition also appears to have played a role in the evolution of *Y. pestis* (42), analysis of *Bordetella* genome sequences (28) and data presented here indicate that little or no gene acquisition has occurred during the divergence of *B. pertussis* from *B. bronchiseptica*.

In the large and diverse collection of strains surveyed here, only 6.4% (248 of 3,849 array elements) of the *B. pertussis* genome was found to be variably present. No known virulence-associated genes were located in RDs, nor were RDs exclusively associated with collections of strains from either before or after vaccination. Although some sets of RDs are missing in almost all the same strains, no two RDs show complete linkage. Instead, the pattern of RDs across the strains has a striking mosaic quality, suggesting that the RDs are not solely vertically inherited. Several nonexclusive factors may account for this observation. Some regions may be hot spots for gene loss, either due to a local feature of the genome (e.g., accessible chromosomal structure or spacing of IS elements) or because the genes in these RDs are nonessential. In this scenario, we would expect that in the absence of new or frequency-dependent selection pressures, all strains eventually would lose all variably present genes. The fact that this state is not observed may indicate that the strains are in an intermediate stage of genome reduction.

Another possibility is that variability in RDs is maintained in the population due to a form of frequency-dependent selection (21), such that an uncommon type of strain is favored due to acquired host immunity against the more common circulating strains. This model assumes that each RD encodes an immunogenic protein, the deletion of which would confer a selective advantage if the acquired ability to evade immunity outweighed the fitness cost of losing gene products in the RD. If the population size was large enough to avoid stochastic extinction, the proportions of various types of strains in the population might fluctuate indefinitely. The mosaic structure of RDs also could be maintained due to independent events of gene loss and interstrain gene transfer, a process that might occur because of a selective advantage conferred by temporary possession of the RDs or as a by-product of the large numbers of IS elements in the genomes.

Homogeneity in gene content among *B. pertussis* strains belies an extraordinarily high amount of genome rearrangement in this species, as first demonstrated by low-resolution chromosomal mapping (32, 33). Homologous recombination between high-copy-number IS481 elements is a likely mechanism for chromosomal rearrangement (28). Using a technique combining PFGE and microarray hybridization, we discovered multiple sites of genome rearrangement in five strains. All but one of the apparent recombination junctions in these strains were adjacent to IS481 elements. However, the rate of IS-mediated rearrangements during natural infection and transfer or lab passage is unknown. We found no changes in gene content of

SwaI fragments in the laboratory strain Tohama I after it was passaged over 200 times on plates or in a recent clinical strain after 12 passages. Most previous investigations have not detected differences in PFGE types (predominantly using XbaI restriction patterns) after a small number of lab passages (2, 15), although one report indicated changes in PFGE types after approximately 10 passages (4). Comparison of PFGE typing of one strain used for production of a Swedish whole-cell vaccine revealed different patterns between vials prepared over an 8-year span, and occasionally within the same vial, but no changes were observed when several isolates were serially subcultured eight times (2). Laboratory conditions may not stimulate or select for these rearrangements to the same degree as natural infection conditions, and so we have initiated investigations into whether rearrangement occurs during chains of transmission of B. pertussis from one host to another.

Although the significance of nonprogrammed chromosomal rearrangement for bacterial evolution has not been heavily studied, some observations suggest that spontaneous chromosomal rearrangement may play a role in the evolution of other bacteria that bear high loads of insertion elements. For example, 9 out of 10 sucrose-resistant Burkholderia mallei mutants had deletions of sacB that were apparently mediated by recombination between flanking IS copies (27). In addition, several authors have proposed that the large number of IS elements in Y. pestis may contribute to rapid adaptive microevolution (34, 42).

Even in bacteria without large numbers of repeated sequences, chromosomal rearrangement has been observed in strains or species that are found only in specific niches, including host-specialized Salmonella species (13) and P. aeruginosa isolates from chronic lung infections in cystic fibrosis patients (17). Furthermore, impaired host colonization of a recombination-deficient Helicobacter pylori mutant strain suggests a functional role for chromosomal rearrangement in H. pylori pathogenesis (29). Like mutator phenotypes, which are observed in some pathogenic species (14, 19, 20), high-frequency chromosomal rearrangement may provide transient variation for rapid adaptation to a new host environment in host-restricted or niche-adapted strains such as B. pertussis.

The high IS load of B. pertussis may reflect a recent proliferation of the repeat elements during an expansion into a new niche that imposed less selective pressure against the concomitant disruption of genes (28), and the perceived "maintenance" of IS elements may be completely nonadaptive. In fact, IS-mediated gene disruption in B. pertussis could be deleterious most of the time, but perhaps the concomitant generation of diversity provided by genome rearrangement is important for the continuing survival of B. pertussis in its ever-changing human niche.

Effects of genome rearrangement might be manifested at the level of transcription, because genes near the rearrangement breakpoints could be put under the control of new transcriptional regulatory elements. Comparison of expression profiles from ancestral and passaged strains of Tohama I and Bpe280 (a recent clinical isolate) revealed a number of differences between the ancestral and passaged isolates, some of which may be due to undetected genome rearrangements or other mutations. For a bacterial species with constrained gene content, such as B. pertussis, alteration of expression patterns is an efficient way to respond to selective pressures. Several reports support the proposition that closely related strains, including those derived by lab passage or growth in a single host, can show significant differences in expression profiles (6, 11, 12, 16, 24, 39). The observation that expression profiles change over relatively few passages has significant implications for the study of B. pertussis, because the lack of a natural animal host for this species requires many investigations of gene regulation and function be conducted in vitro, often with laboratory strains that have been passaged many times. In a broader evolutionary sense, the finding that B. pertussis displays variation in gene expression provides counterbalance to the high level of gene content homogeneity observed in this species and suggests that this pathogen can respond to the dynamic pressures of host immunity by altering regulation of a restricted set of genes, without the benefit of the genetic input that appears to play a significant role in the evolution of other pathogens.

REFERENCES

1. Achtman, M., G. Morelli, P. Zhu, T. Wirth, I. Diehl, B. Kusecek, A. J. Vogler, D. M. Wagner, C. J. Allender, W. R. Easterday, V. Chenal-Francisque, P. Worsham, N. R. Thomson, J. Parkhill, L. E. Lindler, E. Carniel, and P. Keim. 2004. Microevolution and history of the plague bacillus, Yersinia pestis. Proc. Natl. Acad. Sci. USA 101:17837–17842.
2. Advani, A., D. Donnelly, and H. Hallander. 2004. Reference system for characterization of Bordetella pertussis pulsed-field gel electrophoresis profiles. J. Clin. Microbiol. 42:2890–2897.
3. Akerley, B. J., and J. F. Miller. 1993. Flagellin gene transcription in Bordetella bronchiseptica is regulated by the BvgAS virulence control system. J. Bacteriol. 175:3468–3479.
4. Beall, B., P. K. Cassiday, and G. N. Sanden. 1995. Analysis of Bordetella pertussis isolates from an epidemic by pulsed-field gel electrophoresis. J. Clin. Microbiol. 33:3083–3086.
5. Boursaux-Eude, C., S. Thiberge, G. Carletti, and N. Guiso. 1999. Intranasal murine model of Bordetella pertussis infection. II. Sequence variation and protection induced by a tricomponent acellular vaccine. Vaccine 17:2651–2660.
6. Cummings, C. A., H. J. Bootsma, D. A. Relman, and J. F. Miller. 2006. Species- and strain-specific control of a complex, flexible regulon by Bordetella BvgAS. J. Bacteriol. 188:1775–1785.
7. Cummings, C. A., M. M. Brinig, P. W. Lepp, S. van de Pas, and D. A. Relman. 2004. Bordetella species are distinguished by patterns of substantial gene loss and host adaptation. J. Bacteriol. 186:1484–1492.
8. Diavatopoulos, D. A., C. A. Cummings, L. M. Schouls, M. M. Brinig, D. Relman, and F. R. Mooi. 2005. Bordetella pertussis, the causative agent of whooping cough, evolved from a distinct, human-associated lineage of B. bronchiseptica. PLoS Pathog. 1:e45.
9. Dorrell, N., J. A. Mangan, K. G. Laing, J. Hinds, D. Linton, H. Al-Ghusein, B. G. Barrell, J. Parkhill, N. G. Stoker, A. V. Karlyshev, P. D. Butcher, and B. W. Wren. 2001. Whole genome comparison of Campylobacter jejuni human isolates using a low-cost microarray reveals extensive genetic diversity. Genome Res. 11:1706–1715.
10. Fitzgerald, J. R., D. E. Sturdevant, S. M. Mackie, S. R. Gill, and J. M. Musser. 2001. Evolutionary genomics of Staphylococcus aureus: Insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc. Natl. Acad. Sci. USA 98:8821–8826.
11. Gao, Q., K. E. Kripke, A. J. Saldanha, W. Yan, S. Holmes, and P. M. Small. 2005. Gene expression diversity among Mycobacterium tuberculosis clinical isolates. Microbiology 151:5–14.
12. Gaynor, E. C., S. Cawthraw, G. Manning, J. K. MacKichan, S. Falkow, and D. G. Newell. 2004. The genome-sequenced variant of Campylobacter jejuni NCTC 11168 and the original clonal clinical isolate differ markedly in colonization, gene expression, and virulence-associated phenotypes. J. Bacteriol. 186:503–517.

13. **Helm, R. A., A. G. Lee, H. D. Christman, and S. Maloy.** 2003. Genomic rearrangements at *rrn* operons in *Salmonella*. Genetics **165:**951–959.

14. **Hughes, D.** 2000. Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. Genome Biol **1:**REVIEWS0006.1–0006.8.

15. **Khattak, M. N., and R. C. Matthews.** 1993. Genetic relatedness of *Bordetella* species as determined by macrorestriction digests resolved by pulsed-field gel electrophoresis. Int. J. Syst. Bacteriol. **43:**659–664.

16. **Kirisits, M. J., L. Prost, M. Starkey, and M. R. Parsek.** 2005. Characterization of colony morphology variants isolated from *Pseudomonas aeruginosa* biofilms. Appl. Environ Microbiol. **71:**4809–4821.

17. **Kresse, A. U., S. D. Dinesh, K. Larbig, and U. Romling.** 2003. Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs. Mol. Microbiol. **47:**145–158.

18. **Kukkonen, M., M. Suomalainen, P. Kyllonen, K. Lahteenmaki, H. Lang, R. Virkola, I. M. Helander, O. Holst, and T. K. Korhonen.** 2004. Lack of O-antigen is essential for plasminogen activation by *Yersinia pestis* and *Salmonella enterica*. Mol. Microbiol. **51:**215–225.

19. **LeClerc, J. E., and T. A. Cebula.** 2000. *Pseudomonas* survival strategies in cystic fibrosis. Science **289:**391–392.

20. **LeClerc, J. E., B. Li, W. L. Payne, and T. A. Cebula.** 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. Science **274:**1208–1211.

21. **Levin, B. R.** 1988. Frequency-dependent selection in bacterial populations. Philos. Trans. Royal Soc. London B **319:**459–472.

22. **Maurelli, A. T., R. E. Fernandez, C. A. Bloch, C. K. Rode, and A. Fasano.** 1998. "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. Proc. Natl. Acad. Sci. USA **95:**3943–3948.

23. **Middendorf, B., and R. Gross.** 1999. Representational difference analysis identifies a strain-specific LPS biosynthesis locus in *Bordetella* spp. Mol. Gen. Genet. **262:**189–198.

24. **Mongodin, E., J. Finan, M. W. Climo, A. Rosato, S. Gill, and G. L. Archer.** 2003. Microarray transcription analysis of clinical *Staphylococcus aureus* isolates resistant to vancomycin. J. Bacteriol. **185:**4638–4643.

25. **Mooi, F. R., H. Hallander, C. H. Wirsing von Konig, B. Hoet, and N. Guiso.** 2000. Epidemiological typing of *Bordetella pertussis* isolates: recommendations for a standard methodology. Eur. J. Clin. Microbiol. Infect. Dis. **19:**174–181.

26. **Musser, J. M., E. L. Hewlett, M. S. Peppler, and R. K. Selander.** 1986. Genetic diversity and relationships in populations of *Bordetella* spp. J. Bacteriol. **166:**230–237.

27. **Nierman, W. C., D. DeShazer, H. S. Kim, H. Tettelin, K. E. Nelson, T. Feldblyum, R. L. Ulrich, C. M. Ronning, L. M. Brinkac, S. C. Daugherty, T. D. Davidsen, R. T. Deboy, G. Dimitrov, R. J. Dodson, A. S. Durkin, M. L. Gwinn, D. H. Haft, H. Khouri, J. F. Kolonay, R. Madupu, Y. Mohammoud, W. C. Nelson, D. Radune, C. M. Romero, S. Sarria, J. Selengut, C. Shamblin, S. A. Sullivan, O. White, Y. Yu, N. Zafar, L. Zhou, and C. M. Fraser.** 2004. Structural flexibility in the *Burkholderia mallei* genome. Proc. Natl. Acad. Sci. USA **101:**14246–14251.

28. **Parkhill, J., M. Sebaihia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A.** Cronin, P. Davis, J. Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabbinowitsch, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell, and D. J. Maskell. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat. Genet. **35:**32–40.

29. **Robinson, K., M. F. Loughlin, R. Potter, and P. J. Jenks.** 2005. Host adaptation and immune modulation are mediated by homologous recombination in *Helicobacter pylori*. J. Infect. Dis. **191:**579–587.

30. **Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and S. Falkow.** 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. Proc. Natl. Acad. Sci. USA **97:**14668–14673.

31. **Serres, M. H., and M. Riley.** 2000. MultiFun, a multifunctional classification scheme for *Escherichia coli* K-12 gene products. Microb. Comp. Genomics **5:**205–222.

32. **Stibitz, S., and M. S. Yang.** 1997. Genomic fluidity of *Bordetella pertussis* assessed by a new method for chromosomal mapping. J. Bacteriol. **179:**5820–5826.

33. **Stibitz, S., and M. S. Yang.** 1999. Genomic plasticity in natural populations of *Bordetella pertussis*. J. Bacteriol. **181:**5512–5515.

34. **Tong, Z., D. Zhou, Y. Song, L. Zhang, D. Pei, Y. Han, X. Pang, M. Li, B. Cui, J. Wang, Z. Guo, Z. Qi, L. Jin, J. Zhai, Z. Du, X. Wang, J. Yu, P. Huang, H. Yang, and R. Yang.** 2005. Pseudogene accumulation might promote the adaptive microevolution of *Yersinia pestis*. J. Med. Microbiol. **54:**259–268.

35. **Tusher, V. G., R. Tibshirani, and G. Chu.** 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA **98:**5116–5121.

36. **van Boven, M., F. R. Mooi, J. F. Schellekens, H. E. de Melker, and M. Kretzschmar.** 2005. Pathogen adaptation under imperfect vaccination: implications for pertussis. Proc. Biol. Sci. **272:**1617–1624.

37. **van der Zee, A., F. Mooi, J. Van Embden, and J. Musser.** 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.

38. **van Loo, I. H., K. J. Heuvelman, A. J. King, and F. R. Mooi.** 2002. Multilocus sequence typing of *Bordetella pertussis* based on surface protein genes. J. Clin. Microbiol. **40:**1994–2001.

39. **von Gotz, F., S. Haussler, D. Jordan, S. S. Saravanamuthu, D. Wehmhoner, A. Strussmann, J. Lauber, I. Attree, J. Buer, B. Tummler, and I. Steinmetz.** 2004. Expression analysis of a highly adherent and cytotoxic small colony variant of *Pseudomonas aeruginosa* isolated from a lung of a patient with cystic fibrosis. J. Bacteriol. **186:**3837–3847.

40. **Weber, C., C. Boursaux-Eude, G. Coralie, V. Caro, and N. Guiso.** 2001. Polymorphism of *Bordetella pertussis* isolates circulating for the last 10 years in France, where a single effective whole-cell vaccine has been used for more than 30 years. J. Clin. Microbiol. **39:**4396–4403.

41. **Wolfgang, M. C., B. R. Kulasekara, X. Liang, D. Boyd, K. Wu, Q. Yang, C. G. Miyada, and S. Lory.** 2003. Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. USA **100:**8484–8489.

42. **Wren, B. W.** 2003. The yersiniae—a model genus to study the rapid evolution of bacterial pathogens. Nat. Rev. Microbiol. **1:**55–64.