

Complete Genome Sequences of Four *Bordetella pertussis* Vaccine Reference Strains from Serum Institute of India

Michael R. Weigand,^a Yanhui Peng,^a Vladimir Loparev,^a Taccara Johnson,^a Phalasy Juieng,^a Sunil Gairola,^b Rakesh Kumar,^b Umesh Shaligram,^b Ramnath Gowrishankar,^a Hercules Moura,^a Jon Rees,^a David M. Schieltz,^a Yulanda Williamson,^a Adrian Woolfitt,^a John Barr,^a M. Lucia Tondella,^a Margaret M. Williams^a

Centers for Disease Control and Prevention, Atlanta, Georgia, USA^a; Serum Institute of India Pvt. Ltd., Pune, India^b

Serum Institute of India is among the world's largest vaccine producers. Here, we report the complete genome sequences for four *Bordetella pertussis* strains used by Serum Institute of India in the production of whole-cell pertussis vaccines.

Received 21 October 2016 Accepted 26 October 2016 Published 22 December 2016

Citation Weigand MR, Peng Y, Loparev V, Johnson T, Juieng P, Gairola S, Kumar R, Shaligram U, Gowrishankar R, Moura H, Rees J, Schieltz DM, Williamson Y, Woolfitt A, Barr J, Tondella ML, Williams MM. 2016. Complete genome sequences of four *Bordetella pertussis* vaccine reference strains from Serum Institute of India. *Genome Announc* 4(6):e01404-16. doi:10.1128/genomeA.01404-16.

Copyright © 2016 Weigand et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Michael R. Weigand, mweigand@cdc.gov.

Whooping cough (pertussis) is a contagious respiratory disease primarily caused by the Gram-negative bacterium *Bordetella pertussis*. Vaccination programs in many countries have successfully reduced disease incidence. Yet *B. pertussis* persists, and pertussis disease has resurged, in part due to genetic divergence of circulating strains. Although reference strains used for vaccine production vary (1), growing evidence suggests that *B. pertussis* is evolving under vaccine-driven selection (2–5). Here, we report the complete genome sequences of four strains (134, 509, 6229, 25525) used by Serum Institute of India to manufacture whole-cell pertussis vaccines.

Whole-genome shotgun sequencing was performed using a combination of the PacBio RSII (Pacific Biosciences, Menlo Park, CA, USA), Illumina MiSeq (Illumina, San Diego, CA, USA), and Argus (OpGen, Gaithersburg, MD, USA) platforms as described previously (6). Briefly, genomic DNA libraries were prepared for PacBio sequencing using the SMRTbell template prep kit version 1.0 and the polymerase binding kit P6 version 2; libraries for Illumina sequencing were prepared using the NEB ultra library prep kit (New England Biolabs, Ipswich, MA, USA). *De novo* genome assembly was performed using the Hierarchical Genome Assembly Process version 3 (Pacific Biosciences) at >140× coverage.

The resulting consensus sequences were manually checked for circularity and reordered to match the start of Tohama I (CP010964) (6). To ensure accuracy, assemblies were confirmed by comparison to *KpnI* restriction digest optical maps using the Argus system (OpGen) with MapSolver version 2.1.1 (OpGen). For strains 6229 and 25525, putative repeat duplications identified by increased read coverage depth and optical map misalignment were resolved manually. Sequences were further “polished” by mapping Illumina MiSeq PE-300 reads using CLC Genomics Workbench version 9 (CLC bio, Boston, MA, USA). Final assemblies were annotated using NCBI's Prokaryotic Genome Annotation Pipeline.

Isolate and assembly characteristics are summarized in Table 1. All four assemblies included the full complement of known (>40) *B. pertussis* virulence-associated genes. Assembled genomes varied in sequence and chromosomal structure, with 509 appearing similar to vaccine reference strain 10536 (CP012128) (7) and 134 matching a recent sequence of the same strain (CP016338) (7). Genomes of strains 6229 and 25525 were closely related and more similar to clinical isolates than to other vaccine reference strains when compared to available complete assemblies. Two genomes included direct duplication of an approximately 128-kb region flanked by copies of *IS481* that was present in two copies in

TABLE 1 Characteristics of *B. pertussis* vaccine reference strains and genome assemblies

Strain	Genotype ^a	Genome size (bp)	CDSs ^b	Repeats ^c	Accession no.
134	<i>prn1-ptxP1-ptxA2</i>	4,128,984	3,645	NA ^d	CP017402
509	<i>prn7-ptxP2-ptxA4</i>	4,140,370	3,650	NA	CP017403
6229	<i>prn1-ptxP1-ptxA1</i>	4,257,407	3,767	1,324,103 to 1,452,037 1,453,081 to 1,581,015	CP017404
25525	<i>prn1-ptxP1-ptxA1</i>	4,386,396	3,882	1,324,106 to 1,452,040 1,453,084 to 1,581,018 1,582,062 to 1,709,996	CP017405

^a All were *fimH1* and *ptxB2*.

^b CDSs, coding sequences.

^c Coordinates of direct repeats.

^d NA, not applicable.

6229 and three copies in 25525 (Table 1). These duplications were not resolvable by sequencing alone, and proper assembly was achieved only with the aid of optical mapping. Gene content within this region was identical to Tohama I (BP1269 to BP1395, NC_002929) and encoded functions such as amino acid transport, stress responses, and flagellar biosynthesis.

Multiple alignment of complete assemblies has shown that the *B. pertussis* genome exhibits considerable rearrangement plasticity (6, 8, 9) but has thus far not revealed large repeats like those in 6229 and 25525. However, duplication of genes within this same region was inferred by microarray hybridization in Finnish isolate KKK1330 (10). Homologous recombination between copies of IS481 has contributed to genome reduction in *B. pertussis* (11) and these data suggest that expansion is also possible by the same mechanism.

Accession number(s). The complete genome sequences have been deposited at DDBJ/EMBL/GenBank under the accession numbers listed in Table 1. The versions described in this paper are the first versions.

ACKNOWLEDGMENTS

We thank Pam Cassidy for technical assistance with bacterial culture. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

FUNDING INFORMATION

This work was supported by internal funds.

REFERENCES

- Kilgore PE, Salim AM, Zervos MJ, Schmitt HJ. 2016. Pertussis: microbiology, disease, treatment, and prevention. *Clin Microbiol Rev* 29: 449–486. <http://dx.doi.org/10.1128/CMR.00083-15>.
- Bart MJ, Harris SR, Advani A, Arakawa Y, Bottero D, Bouchez V, Cassidy PK, Chiang CS, Dalby T, Fry NK, Gaillard ME, van Gent M, Guiso N, Hallander HO, Harvill ET, He Q, van der Heide HG, Heuvelman K, Hozbor DF, Kamachi K, Karataev GI, Lan R, Lutyńska A, Maharjan RP, Mertsola J, Miyamura T, Octavia S, Preston A, Quail MA, Sintchenko V, Stefanelli P, Tondella ML, Tsang RS, Xu Y, Yao SM, Zhang S, Parkhill J, Mooi FR. 2014. Global population structure and evolution of *Bordetella pertussis* and their relationship with vaccination. *mBio* 5:e01074. <http://dx.doi.org/10.1128/mBio.01074-14>.
- Octavia S, Maharjan RP, Sintchenko V, Stevenson G, Reeves PR, Gilbert GL, Lan R. 2011. Insight into evolution of *Bordetella pertussis* from comparative genomic analysis: evidence of vaccine-driven selection. *Mol Biol Evol* 28:707–715. <http://dx.doi.org/10.1093/molbev/msq245>.
- Sealey KL, Harris SR, Fry NK, Hurst LD, Gorringer AR, Parkhill J, Preston A. 2015. Genomic analysis of isolates from the United Kingdom 2012 pertussis outbreak reveals that vaccine antigen genes are unusually fast evolving. *J Infect Dis* 212:294–301. <http://dx.doi.org/10.1093/infdis/jiu665>.
- Xu Y, Liu B, Grondahl-Yli-Hannuksila K, Tan Y, Feng L, Kallonen T, Wang L, Peng D, He Q, Wang L, Zhang S. 2015. Whole-genome sequencing reveals the effect of vaccination on the evolution of *Bordetella pertussis*. *Sci Rep* 5:12888.
- Bowden KE, Weigand MR, Peng Y, Cassidy PK, Sammons S, Knipe K, Rowe LA, Loparev V, Sheth M, Weening K, Tondella ML, Williams MM. 2016. Genome structural diversity among 31 *Bordetella pertussis* isolates from two recent U.S. whooping cough statewide epidemics. *mSphere* 1:e00036-16.
- Weigand MR, Peng Y, Loparev V, Batra D, Burroughs M, Johnson T, Juieng P, Rowe LA, Tondella ML, Williams MM. 2016. Complete genome sequences of *Bordetella pertussis* vaccine reference strains 134 and 10536. *Genome Announc* 4(5):e00979-16. <http://dx.doi.org/10.1128/genomeA.00979-16>.
- Bart MJ, van der Heide HG, Zeddeman A, Heuvelman K, van Gent M, Mooi FR. 2015. Complete genome sequences of 11 *Bordetella pertussis* strains representing the pandemic *ptxP3* lineage. *Genome Announc* 3(6): e01394-15. <http://dx.doi.org/10.1128/genomeA.01394-15>.
- Belcher T, Preston A. 2015. *Bordetella pertussis* evolution in the (functional) genomics era. *Pathog Dis* 73:ftv064. <http://dx.doi.org/10.1093/femspd/ftv064>.
- Heikkinen E, Kallonen T, Saarinen L, Sara R, King AJ, Mooi FR, Soini JT, Mertsola J, He Q. 2007. Comparative genomics of *Bordetella pertussis* reveals progressive gene loss in Finnish strains. *PLoS One* 2:e904. <http://dx.doi.org/10.1371/journal.pone.0000904>.
- Brinig MM, Cummings CA, Sanden GN, Stefanelli P, Lawrence A, Relman DA. 2006. Significant gene order and expression differences in *Bordetella pertussis* despite limited gene content variation. *J Bacteriol* 188: 2375–2382. <http://dx.doi.org/10.1128/JB.188.7.2375-2382.2006>.