

James Madison University

JMU Scholarly Commons

Senior Honors Projects, 2010-2019

Honors College

Summer 2018

Evolution of bordetella pertussis genome may play a role in the increased rate of whooping cough cases in the United States

Kevin Loftus

James Madison University

Follow this and additional works at: <https://commons.lib.jmu.edu/honors201019>



Part of the [Bacteriology Commons](#), [Bioinformatics Commons](#), [Biotechnology Commons](#), [Community Health and Preventive Medicine Commons](#), [Epidemiology Commons](#), [Genetic Structures Commons](#), [Immunology of Infectious Disease Commons](#), and the [Pathogenic Microbiology Commons](#)

Recommended Citation

Loftus, Kevin, "Evolution of bordetella pertussis genome may play a role in the increased rate of whooping cough cases in the United States" (2018). *Senior Honors Projects, 2010-2019*. 522.

<https://commons.lib.jmu.edu/honors201019/522>

This Thesis is brought to you for free and open access by the Honors College at JMU Scholarly Commons. It has been accepted for inclusion in Senior Honors Projects, 2010-2019 by an authorized administrator of JMU Scholarly Commons. For more information, please contact dc_admin@jmu.edu.

Evolution of *Bordetella pertussis* genome may play a role in the increased rate of whooping cough cases in the United States

An Honors College Project

by Kevin Loftus

James Madison University

FACULTY COMMITTEE:

HONORS COLLEGE APPROVAL:

Project Advisor: Dr. Louise Temple
Department of Integrated Science and Technology

Bradley R. Newcomer, Ph.D.,
Dean, Honors College

Reader: Dr. Amanda Biesecker
Department of Integrated Science and Technology

Reader: Dr. Steven Cresawn
Biology Department

PUBLIC PRESENTATION

This work was presented at Bordetella Research Day in Baltimore Maryland on April 6th 2018 and the ISAT Senior Symposium in Harrisonburg Virginia on April 20th 2018.

Abstract

Bordetella pertussis is the bacterium responsible for pertussis, a disease commonly referred to as whooping cough. Recently, pertussis has made a resurgence in the U.S. despite high-vaccination coverage. Possible causes of the increased number of pertussis cases include genetic evolution of *B. pertussis*, increased awareness of the disease, better laboratory diagnostics, and the switch from a whole-cellular (wP) vaccine to an acellular vaccine (aP) in the 1990s. Fortunately, just as *B. pertussis* is evolving, so is the arsenal of technologies used to understand and combat this pathogenic bacterium. Whole genome sequencing is one technology that helps researchers better understand the evolution *B. pertussis*. This project included the isolation of genomic DNA and sequencing of two novel *B. pertussis* strains isolated from patients in Virginia. This project also utilized bioinformatics to analyze data obtained from genomes of twelve previously sequenced *B. pertussis* isolates. Specifically, the genomes were analyzed and compared to each other and to vaccine reference strains. PubMLST, a database hosted by the University of Oxford, was used to perform multi-locus sequence typing (MLST) and to determine alleles of genes that encode for pertussis toxin subunit 1 (*ptxA*), the pertussis toxin promoter (*ptxP*), pertactin (*prn*), and fimbriae serotype 3 (*fim3*), which are common components in acellular pertussis vaccines. Protein variants were then determined by comparing the protein sequences of PtxA, Prn, and Fim3 to known references using NCBI BLAST. Data gathered from PubMLST and protein alignments support other studies that indicate *B. pertussis* may be evolving to evade pertussis vaccines. Notably, all of the 14 strains analyzed in this paper carry a *ptxP3* allele whereas vaccine reference strains such as Tohama 1 and C393 carry a *ptxP1* allele. The other genes analyzed, *ptxA*, *prn*, and *fim3*, also showed divergence from the allele types of Tohama 1 and C393. This project builds on information gathered from previous studies and contributes to the growing knowledge of the evolution of the *B. pertussis* genome as related to whooping cough.

TABLE OF CONTENTS

I.	Introduction.....	3
	a. Problem overview.....	3
	b. The culprit, <i>Bordetella pertussis</i>	4
	i. Structure and virulence factors.....	4
	ii. Transmission and pathogenesis.....	5
	a. Symptoms, diagnosis, and prevention.....	7
	iii. Symptoms.....	7
	iv. Diagnosis.....	8
	v. Prevention.....	10
	b. Public policy, cultural dynamics and stakeholders.....	11
	vi. Public policy and cultural dynamics.....	11
	vii. Stakeholders analysis.....	13
II.	Objectives.....	14
	a. Contribute to the growing knowledge of the evolution of the <i>B. pertussis</i> genome....	14
	i. Perform DNA isolation and sequencing of two VA clinical isolates.....	14
	ii. Perform bioinformatics analysis on new strains and twelve previously sequenced VA clinical isolates	15
III.	Materials and methods.....	15
	a. Isolate collection.....	15
	b. DNA purification, sequencing and assembly.....	15
	i. DNA purification.....	15
	ii. Sequencing and assembly.....	16
	c. Genome annotation and annotation comparison.....	16
	i. Genome annotation.....	16
	ii. Genome annotation comparison.....	17
	d. Structural similarity tree construction.....	18
	i. Permutations output file and data formatting.....	18
	ii. Submission of job to MLGO and creation of tree using R.....	18
	e. MLST and virulence factor allele typing.....	18
	i. Overview of BIGSDB and PubMLST.....	18
	ii. MLST and sequence typing.....	19
	iii. Virulence factor allele typing.....	19
IV.	Results and Discussion.....	20
	a. ProgressiveMauve whole-genome alignment.....	20
	b. Cladogram based on LCB order	21
	c. Multi-locus sequence typing (MLST).....	22
	d. Virulence factor allele typing.....	23
	i. Pertactin.....	24
	ii. Pertussis toxin subunit 1 (<i>ptxA</i>).....	25
	iii. Pertussis toxin promoter (<i>ptxP</i>).....	26
	iv. Fimbriae 3 (<i>fim 3</i>).....	26
V.	Conclusions and future work.....	27
VI.	Acknowledgements and funding	29

I. Introduction

I-a Problem overview

Pertussis, also known as whooping cough, is a contagious respiratory disease that is caused primarily by the bacterium *Bordetella pertussis*. Pertussis was first mentioned in literature in England during the mid-sixteenth century and the first epidemic was reported in Paris in 1578 (1, 2) The name pertussis, meaning intense cough in Latin, was given to the illness in the late seventeenth century (2). *B. pertussis* was first isolated in 1906 by French scientists Bordet and Gengou (3). Pertussis outbreaks were a major cause of death, especially for young children, prior to the development of vaccines (4). After the introduction of pertussis vaccines in the 1940s, the number of cases and deaths from pertussis was reduced by around 90% (5). However, despite successful vaccination programs and high vaccination coverage, *B. pertussis* was never eradicated and continues to infect both vaccinated and unvaccinated people. Pertussis is one of the most poorly controlled vaccine-preventable diseases in the United States and around the globe. Recently there has been an increased number of reported pertussis cases in the United States (6). There are several proposed causes of this resurgence which include genetic evolution of *B. pertussis*, increased awareness of the disease, newer laboratory diagnostic techniques, and the switch from a whole-cellular vaccine to an acellular vaccine (7).

Fortunately, just as *B. pertussis* is evolving, so is our arsenal of technologies used to research and combat the bacteria. The emerging field of bioinformatics, which combines biology, computer science, and mathematics, is one way that researchers are using technology to study *B. pertussis*. The ability to sequence whole genomes quickly and affordably with next-generation sequencing (NGS) has greatly increased our knowledge about the bacteria. NGS and bioinformatics can be used to assemble, annotate, and compare genomes of various emerging strains of *B. pertussis* (8). In order to perform a bioinformatics analysis focused on molecular epidemiology it is important to understand the bacterium and illness it causes.

The overall purpose of this project was to utilize NGS data and bioinformatics to analyze several *B. pertussis* genomes that were cultured from isolates from whooping cough patients in Virginia.

One specific goal of this project is to perform a comparative bioinformatics analysis on recent Virginia *B. pertussis* isolates in order to assess the genomic variation. Another specific goal of this project is to build on information gathered from previous studies in an attempt to better understand the *B. pertussis* genome.

I-b The culprit, *Bordetella pertussis*

I-b-i Structure and virulence factors

B. pertussis is a small, aerobic gram-negative bacterial pathogen that infects only humans. Gram-negative bacteria have a lipopolysaccharide (LPS) layer that activates the immune response of their host (9). *B. pertussis* genomes contain mobile IS481 insertion elements (24). IS481 are mobile genetic elements that are thought to play a role in rearrangements and deletions by homologous recombination in the *B. pertussis* genome (24). *B. pertussis* produces many virulence factors, molecules produced by a bacterial pathogen that increase the bacteria's ability to infect hosts and cause disease. This study will focus on the genes that encode for pertussis toxin (Ptx), Fimbriae (Fim), and pertactin (Prn), which are protein components used in aP vaccines.

Pertussis toxin (Ptx) is an exotoxin that utilizes a type IV secretion system (10). This secretion system is encoded by the *ptl* genes and transports Ptx across the bacterial outer membrane (11). Ptx is the cause of leukocytosis and lymphocytosis and is a potential cause of death from pertussis (9). The pertussis toxin promoter (*ptxP*) and pertussis toxin subunit 1 (*ptxA*) will be analyzed in this study. Fim is another virulence factor of *B. pertussis* which are found on the surface of the bacteria and are thought to function as adhesion factors (12, 13). Fimbria type 3 (*fim3*) will be analyzed in this study because it encodes for Fim3, a component in aP vaccines (12). Prn is a large protein, ~70kDa, that is used in aP vaccines. Interestingly, most current circulating strains do not express Prn indicating that it may not be essential to the pathogenesis of current strains (10). These virulence factors, along with other key virulence factors that will not be analyzed in this study, can be seen below in figure 1.

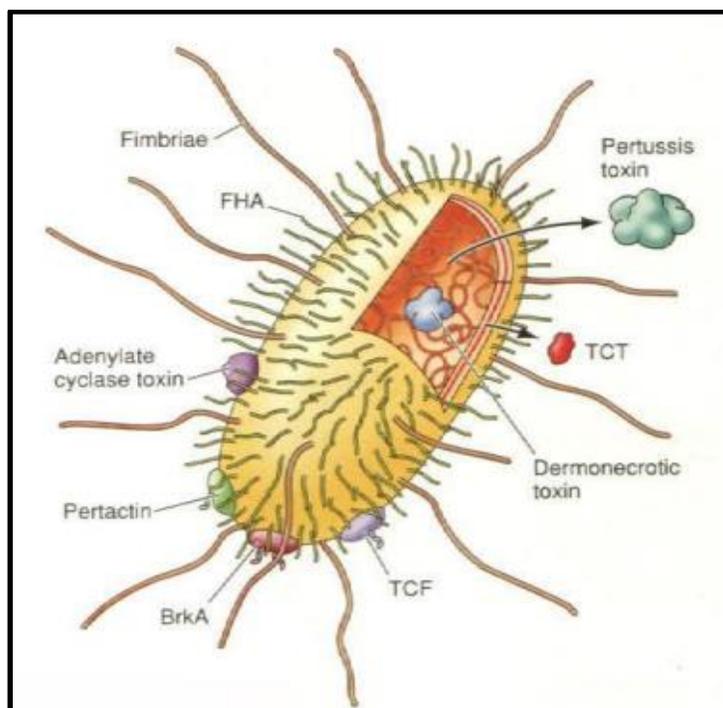


Figure 1. Cartoon of *B. pertussis* showing virulence factors.

<http://www.asm.org/ccLibraryFiles/FILENAME/000000004506/Jan97WeissFeature.pdf>

The virulence factors shown above are what help *B. pertussis* infect a host and cause disease. Protection from infection by *B. pertussis* depends on protective antibodies against bacterial proteins, including these virulence factors (6).

I-b-ii Transmission and pathogenesis

B. pertussis only infects humans and is transmitted by respiratory droplets from the coughing or sneezing of an infected person, making it very contagious (13). A single infected person can spread the disease to as many as 17 other individuals in unprotected populations (14–16). Young children, especially infants under one year old, are particularly at risk to contract pertussis. Infants not only make up the highest percentage of pertussis cases, but are also much more likely to die from complications of the disease (1). Recently, the rate of pertussis cases among adolescents and adults has shown to be increasing (6). While the mortality rates among adolescents and adults remains relatively low, the rise in adolescent and adult cases is still alarming because the vast majority of

pertussis cases in infants can be attributed to transmission from an older family member such as a sibling, parent, or grandparent (9, 17, 18)

The entire mechanism of pathogenesis of *B. pertussis* to humans is not yet clear. The infection process begins in the respiratory tract where *B. pertussis* attaches to the cilia of epithelial cells (19). Once attached to the epithelial cells *B. pertussis* produces toxins that paralyze cilia and inflame the respiratory tract. This inhibits the expulsion of pulmonary secretions which can lead to pulmonary hypertension(20, 21). The antigens created by pertussis seem to assist in the bacterial resistance to the host immune defense. A diagram of the mechanisms of pathogenesis of *B. pertussis* can be seen below in Figure 2.

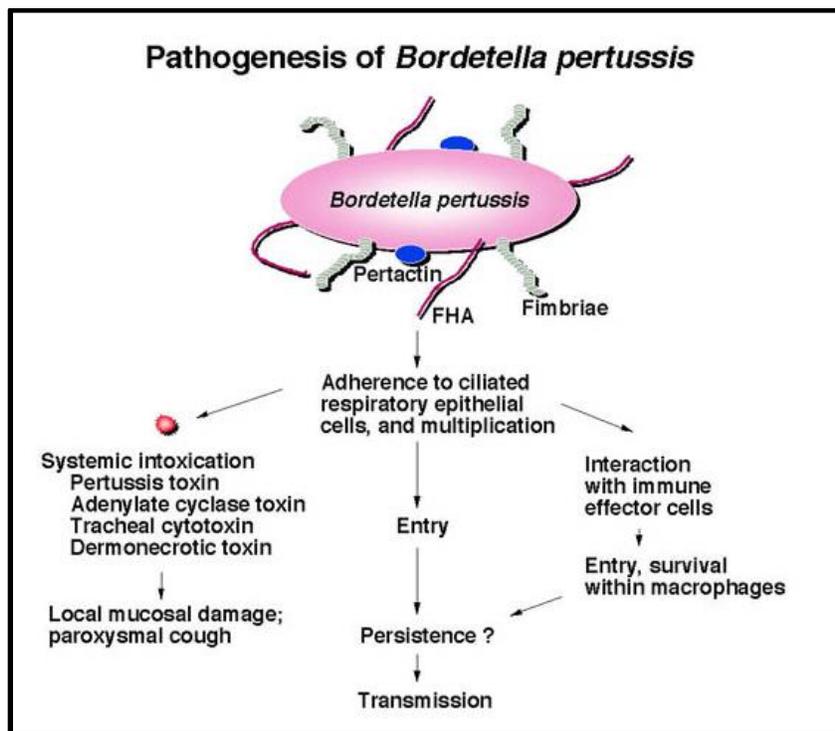


Figure 2. Pathogenesis of *B. pertussis* <https://www.my-pharm.ac.jp/~yishibas/research/Pertussis1.jpg> The process begins with the adherence to epithelial cells and ends with transmission to another host.

I-c Symptoms, diagnosis, and prevention

I-c-i Symptoms

Once *B. pertussis* enters a viable human host the average incubation period ranges from 7-10 days (6). In unimmunized children, the disease is described by three different stages—the catarrhal stage, the paroxysmal stage, and the convalescent stage. During the catarrhal stage, which usually last 1-2 weeks, symptoms can include a mild cough, sneezing, nasal congestion, and a mild sore throat (6). Pertussis is often over looked as a cause for these symptoms because they are similar to the symptoms of a common cold or other infections (9). This is unfortunate because the bacterium is very contagious during this catarrhal stage (6).

The next stage that occurs is the paroxysmal stage. A defining symptom of this phase is the onset of a series of repetitive violent coughs, or paroxysms, in an attempt to clear mucus. After the paroxysm, the patient often inhales deeply then lets out a high-pitched whoop (6). This whoop is a defining characteristic of whooping cough. That is why the disease is usually identifiable as pertussis during this stage. The paroxysms are often more frequent and severe at night and are often followed by vomiting (9). The duration of this stage can last anywhere from 1-6 weeks.

The final stage of the disease is convalescent stage. This stage can last for as little as a week to over several months. During this stage the severity and frequency of the paroxysms are gradually slowed until they stop. However, if the patient contracts a viral respiratory infection, the paroxysms often return (6). An image depicting the timeline of a pertussis infection can be seen below in Figure 3. These stages are typical in unimmunized children, but they can also occur in unprotected or partially protected adolescents and adults. However, the symptoms of the disease in adolescents and adults vary greatly between individuals, sometimes patients are asymptomatic (6). This can be dangerous as an asymptomatic adult can transmit the disease to an infant without even knowing they were infected.

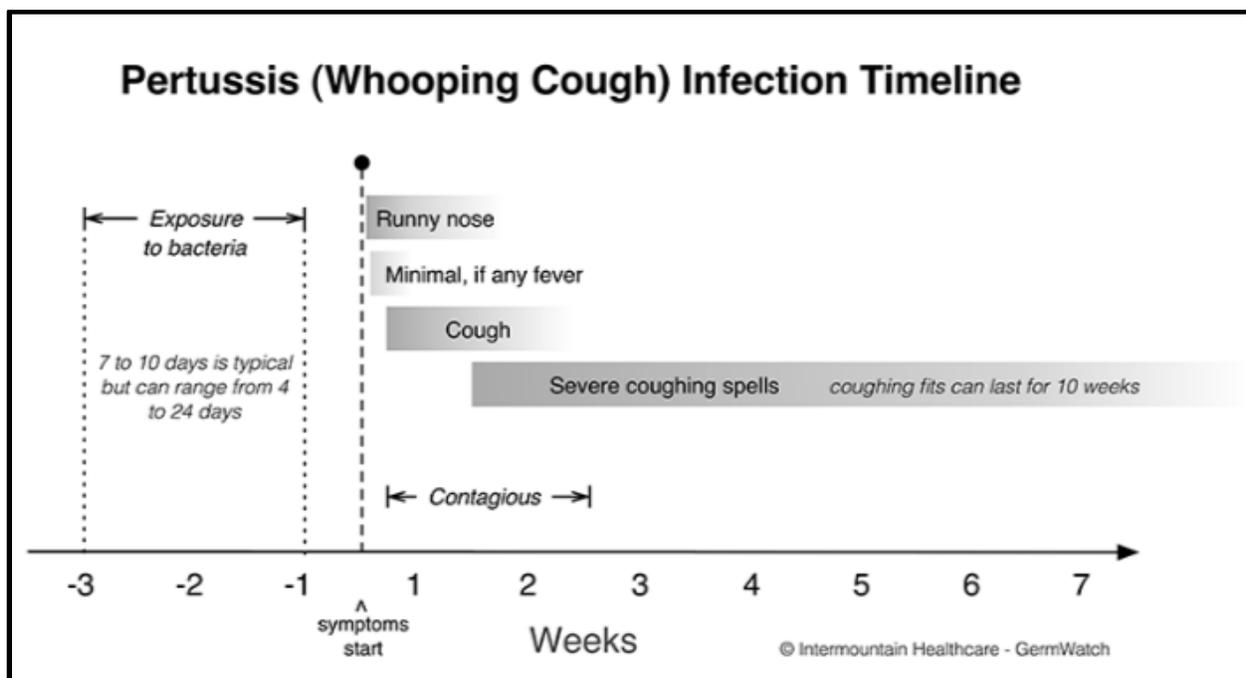


Figure 3. Figure depicting the symptom timeline of a typical pertussis infection

<https://intermountainhealthcare.org/health-information/germwatch/germ-school/bordetella-pertussis/>

There are a number of complications associated with a *B. pertussis* infection. Secondary bacterial pneumonia is not only the most common complication, but also the complication responsible for most pertussis-related deaths (6). The paroxysm coughing associated with whooping cough can also cause seizures and other neurological complications due to a reduced oxygen supply to the brain. Other complications include but are not limited to dehydration, hernias, difficulty sleeping, and rib fracture (6).

I-c-ii Diagnosis

Pertussis is often hard to diagnosis because its early symptoms are similar to other common respiratory illnesses like the common cold. Because healthcare providers often fail to diagnose effectively, it is believed that pertussis is a greater burden to society than the statistics reflect (22). One study conducted on 61 pertussis patients concluded that pertussis was initially misdiagnosed as something else in over 50% of patients (23).

Pertussis can be diagnosed in several ways. The first indicator that might make a clinician test for pertussis is the paroxysmal cough that is unique to pertussis. However, patients sometimes do not cough during a short clinical visit therefore this can be missed by a patient's healthcare provider (22). If a clinician suspects that a patient's symptoms are due to pertussis, culture, polymerase chain reaction (PCR), and serology are the three main techniques used to test patients and confirm a diagnosis.

One way to test for pertussis is with PCR. One advantage of PCR is that it takes much less time than a culture. However, PCR tests are only effective in the beginning of the illness as the bacterial DNA diminishes after the first few weeks of the cough onset (6). Some PCR tests are very sensitive which can lead to a higher rate of false-positive results than with culture (6). Serological tests are another diagnostic tool that are useful after the patient has had the cough for several weeks. The CDC and FDA have developed serological assays that have been successful in diagnosing pertussis (6).

A third way pertussis can be diagnosed is through culture. The culture is usually obtained from the nasopharynx. Cultures are more likely to be positive in the first stages of the disease (6). Cultures can take up to two weeks to give a definitive diagnosis (6). This means that sometimes the results of the culture are received too late to be useful to the patient (6).

The usefulness of each test depends on several factors such as how long the patient has had a cough, the age of the patient, and the history of pertussis cases in the area where the patient lives (6). However, culture has distinct advantages over the other methods. While culturing the bacteria may be more difficult and time consuming than PCR or serological tests, it is essential for comparative genomics. It is only through culturing that circulating strains can be isolated and sequenced for analysis. A graphical depiction of which diagnostic test should be used in relation to how long the patient has had a cough can be seen below in Figure 4.

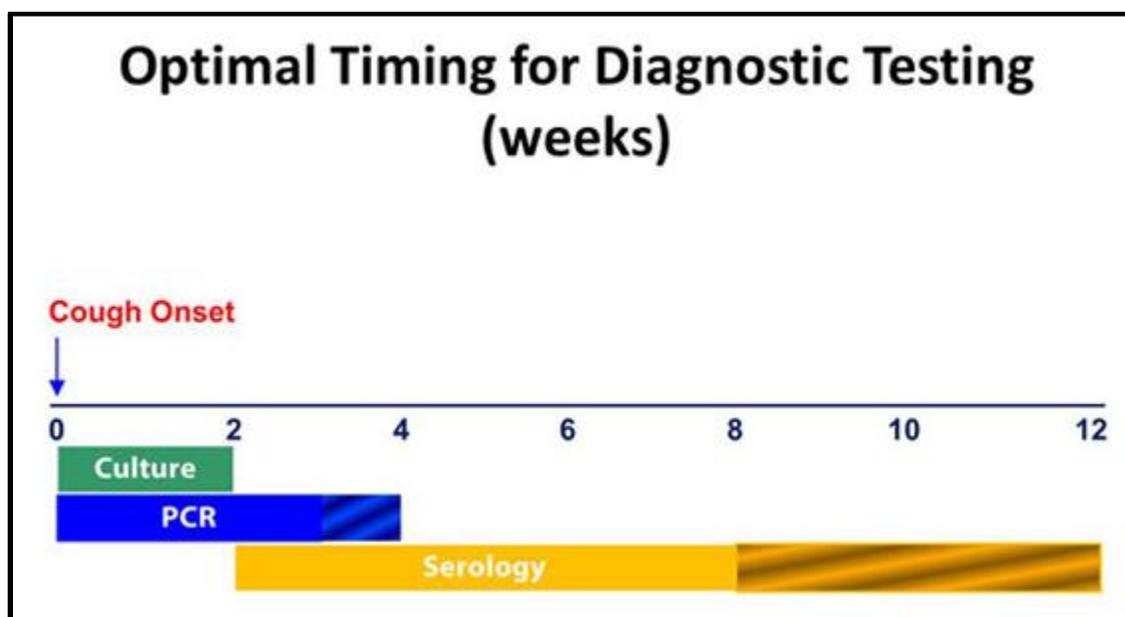


Figure 4. Figure illustrating optimal timeframe for the three laboratory pertussis diagnosis techniques <https://www.cdc.gov/pertussis/clinical/diagnostic-testing/diagnosis-confirmation.html>

This figure shows the optimal test to use in relation to the number of weeks the patient has had a cough.

I-c-iii Prevention

A whole-cell pertussis vaccine consisting of a suspension of inactivated *B. pertussis* was developed in the 1930s. In the 1940s, the vaccine was combined with diphtheria and tetanus toxoids to create a multi-purpose vaccine known as DTP (6). It is estimated that this vaccine reduced the number of pertussis cases by around 90% (5). However, this vaccine had several short-term side effects such as fever and/or pain, swelling, and redness at the injection site (6). These effects caused some people to believe the vaccine was unnecessarily reactogenic. This safety concern led to the development and implementation of less reactogenic acellular vaccine in the 1990s (6).

The acellular vaccine for infants introduced in the 1990s, known as DTaP, consists of proteins instead of a whole-inactivated cell. In 2005, an acellular pertussis, diphtheria, and tetanus vaccine known as Tdap which was designed for adolescents and adults became available (9). Common proteins found in these acellular vaccines are PT, FHA, FIM 1, FIM 3, and PRN.

While these acellular vaccines are less reactogenic, it is widely believed that the switch from a whole-cell to an acellular vaccine is at least partially responsible for the increasing number of pertussis cases (6). An image supporting this hypothesis can be seen below in Figure 5. Some research suggests the acellular vaccine is not only less effective, but also effective for a shorter amount of time than its whole-cell counterparts (6).

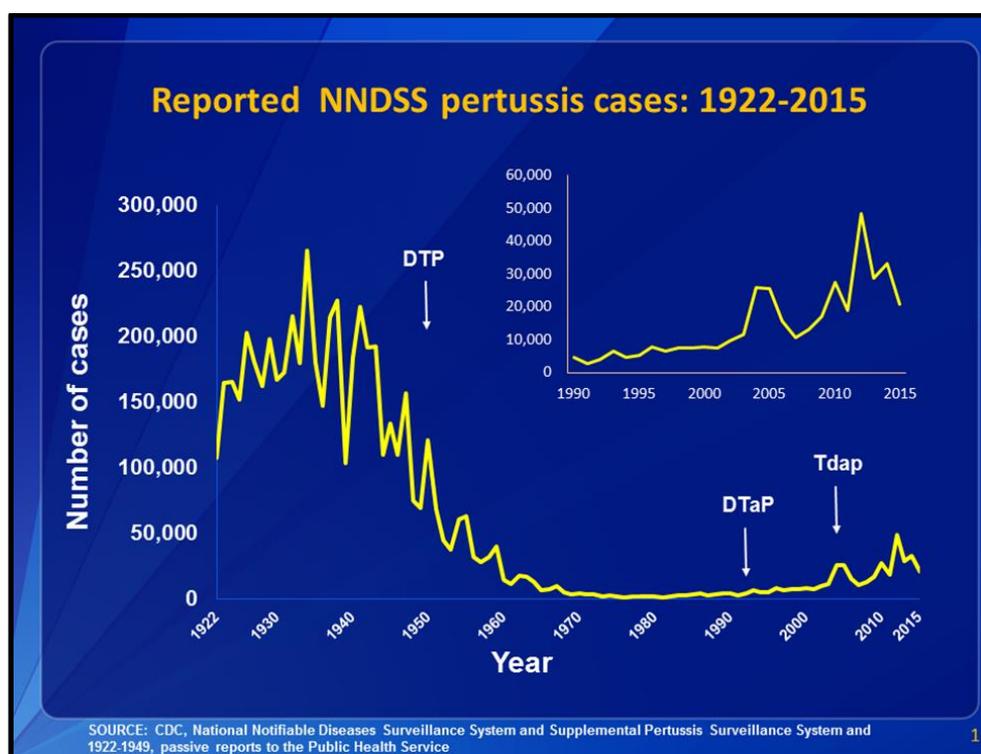


Figure 5. Graph depicting U.S. pertussis cases in relation to time. The introduction date of the various vaccines is depicted on the graph. A sharp drop in pertussis cases is seen with the introduction of the DTP. The recent resurgence is seen in the inlay.

<https://www.cdc.gov/pertussis/images/incidence-graph.png>

I-d Public policy, cultural dynamics and stakeholders

I-d-i Public policy and cultural dynamics

Public policy and cultural dynamics are intertwined in the case of pertussis. As seen in the previous section, vaccination policies are often shaped by public concern of vaccination risk (6). The current U.S. vaccination strategy was developed in response to the public's distrust of the whole-cellular vaccine (24). However, some countries that did not experience similar public outrage still successfully use the whole-cell vaccine (24). The United States encourages vaccination through

compulsory school vaccinations. This means that unless a child has a valid medical or religious reason not to be vaccinated, that child must be vaccinated in order to attend school (24).

These mandates have allowed the United States to achieve vaccination coverage of over 95% (37). This high-vaccination coverage is important in maintaining herd immunity in the United States. The concept of herd immunity states that the higher the vaccination coverage of a population then the less likely an outbreak is to occur. For this reason, when parents decide not to vaccinate their own children, they are not only jeopardizing their children's health, but also may be jeopardizing the health of other children as well. The concept of herd immunity is illustrated in Figure 6 below.

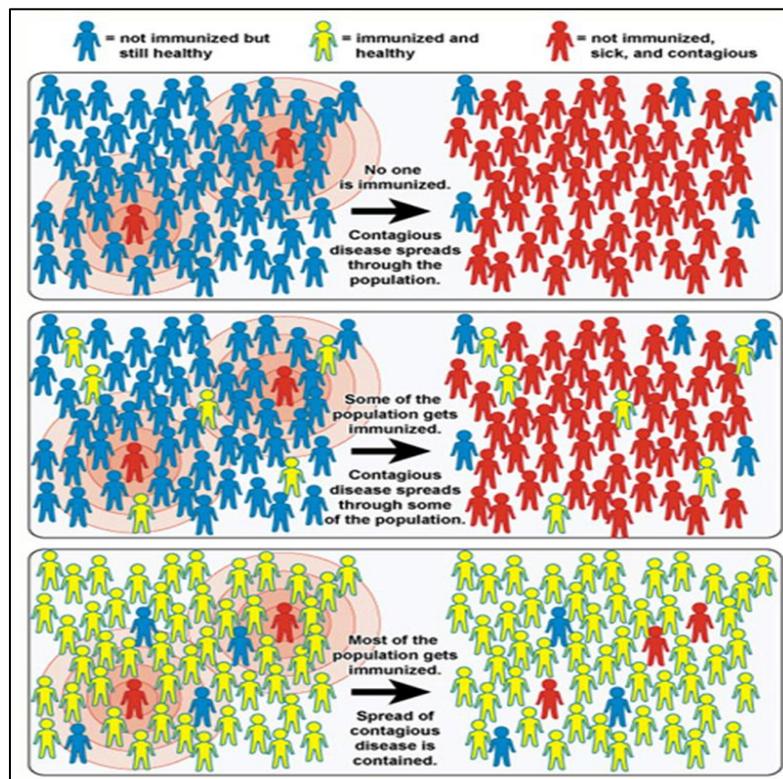


Figure 6. Figure depicting the concept of herd immunity

<http://www.pbs.org/wgbh/nova/body/herd-immunity.html>

I-d-ii Stakeholders analysis

There are many government and private stakeholders involved in developing and improving the United States pertussis vaccine. Stakeholders are involved in a dynamic relationship in order to create vaccines that are effective, safe, and accessible. One group of stakeholders is the general public. The public plays a crucial role in vaccine development by participating in trials (25). Public opinion and concern also plays a major role in developing policy (24).

Another stakeholder in vaccine development is the media. Media plays a large role in shaping public opinion. It is important for the media to make sure they relay accurate information to the general public. If the media provides the public with inaccurate information, it can negatively impact people who trust media as their source of vaccination information (25).

A third group of stakeholders is vaccine manufacturers. Creating a vaccine is a very expensive process with a high chance for failure (25). However, vaccine manufacturing can be a very lucrative business. Top vaccine manufacturers such as Merck and Pfizer generate billions of dollars in revenue each year (25). It is the ethical responsibility of these vaccine manufacturers to develop the safest, and most effective vaccine possible for the public.

Health professionals also play an important role in pertussis treatment and vaccine development. Doctors play a critical role in conveying the importance of vaccinations to their patients and also must be trained to correctly diagnose the disease in their patients (24). Medical researchers also play an important role in leading research for vaccine development.

Another group of stakeholders is government agencies. Government agencies like the FDA make sure vaccine manufacturers are creating safe vaccines. Agencies such as the CDC play an important role in tracking pertussis epidemics as well as conducting important research to understand the disease (24). The government also spends billions of dollars treating vaccine-preventable diseases each year (26). This tax burden could be reduced by the development of more effective vaccines.

While some stakeholders are excluded from this analysis, the final stakeholder mentioned in this analysis is academia. Academia plays an important role in conducting research as well as developing future leaders in various other sectors of stakeholders (25). It is important to note that one person may play the role of many different stakeholders. An example of this would be a physician that may play a role as a clinician, researcher, volunteer, and a parent.

II- Objectives

II-a Contribute to the growing knowledge of the evolution of the *B. pertussis* genome.

The goal of this project is to contribute to the growing knowledge of the *B. pertussis* genome.

II-a-i Perform DNA isolation and sequencing of two VA clinical isolates.

DNA purification was performed on two VA clinical isolates at the University of Virginia's Hewlett Lab. Sequencing was then performed at the Delaware Biotechnology Institute at the University of Delaware. These strains contribute to the growing number of *B. pertussis* whole-genome sequences.

II-a-ii Perform bioinformatics analysis on new strains and twelve previously sequenced VA clinical isolates.

To understand the genomic and proteomic content of the two new strains and twelve previously sequenced VA clinical isolates, a bioinformatics analysis was performed. This included an alignment of the protein sequences using progressiveMauve, the creation of a structural similarity tree using Maximum Likelihood for Gene Order Analysis, Multi-Locus Sequence Typing, and allele typing of key virulence factors.

III- Materials and methods

III-a Isolate Collection

The isolates used in this study were collected by the Virginia Division of Consolidated Laboratory Services (DCLS) from 2001-2015 (27). The isolates are accompanied with patient vaccination status and other clinico-epidemiological data from the Virginia Department of Health (VDH) (27). Isolates accompanied with this information are rare due to a shift from bacterial culture to PCR as the preferred diagnostic method in some regions (27).

III-b DNA purification, sequencing and assembly

III-b-i DNA purification.

To contribute to the increasing number of *B. pertussis* whole genome sequences, two new VA clinical isolates from UVA's repository were purified and sequenced. DNA was purified using the Wizard® Genomic DNA Purification Kit for Isolation of Genomic DNA from Gram Negative Bacteria with modifications. The modifications were made by Dr. Laure Gonyar to optimize genomic based on previous experience. The protocol from Wizard® can be seen below.

The cells were pelleted by centrifuging 1ml of overnight culture for two minutes at 16,000 x g and then discarding the supernatant. The first step in the lysing procedure was to add of 600 µl Nuclei Lysis solution and gently mix by pipetting. The mixture was then incubated for 5 minutes at 80°C and cooled to room temperature. Next 3µl of RNase Solution was added, incubated at 37°C for 45–60 minutes, then cooled to room temperature. Proteins precipitation was done by adding 200µl of Protein Precipitation Solution, vortexing for 30 seconds, incubating on ice for 1 hour, and centrifuging at 16,000 x g for 10 minutes.

The first step in DNA precipitation and rehydration was to transfer the supernatant to a clean tube with 600µl of room temperature isopropanol and mix. The solution was then centrifuged for 20 minutes at 16,000 x g and the supernatant was decanted. We then added 600µl of room temperature 70% ethanol, mixed, and centrifuged for 5 minutes at 16,000 x g. The ethanol was then aspirated, and the pellet was air-dried for 15 minutes. The DNA pellet was rehydrated in

100µl of Rehydration Solution overnight at overnight at 4°C.

After purification, the purity of the DNA from each sample was measured using a nanodrop spectrophotometer to ensure that the concentration of genomic DNA was high enough for sequencing.

III-b-ii Sequencing and assembly

Sequencing was performed by the Delaware Biotechnology Institute at the University of Delaware using Pacific Bioscience's PacBio RSII system. The RSII system utilizes Single-Molecule, Real Time (SMRT) technology (28). SMRT technology works by immobilizing DNA polymerase and template at the bottom of a well and utilizing phospholinked nucleotides to view a DNA strand as it is produced by DNA polymerase (28). First, genomic DNA was used to create SMRTbell template libraries according to PacBio's standard 20-kb library (29). One SMRT cell was used for each strain. Next, sequencing was performed on the PacBio RSII sequencer. Finally, the genomes were assembled *de novo* using the Hierarchical Genome Assembly Process (HGAP)(28). The PacBio RSII sequencer is particularly useful for sequencing of *B. pertussis* because it utilizes long-reads which can improve the accuracy of sequencing genomes with mobile elements and structural rearrangements that are characteristic of the *B. pertussis* genome

III-c Genome annotation and annotation comparison.

III-c-i Genome annotation

Rapid annotation using subsystem technology (RAST) was used to annotate the genomes (30). Genomes were uploaded to RAST and then annotated according to the parameters shown in figure 7 below.

RAST Annotation Settings:	
Choose RAST annotation scheme	Classic RAST ▼
Select gene caller	RAST ▼
Select FIGfam version for this run	Release70 ▼
Automatically fix errors?	<input checked="" type="checkbox"/> Yes
Fix frameshifts?	<input type="checkbox"/> Yes
Build metabolic model?	<input type="checkbox"/> Yes
Backfill gaps?	<input checked="" type="checkbox"/> Yes
Turn on debug?	<input type="checkbox"/> Yes
Set verbose level	0
Disable replication	<input type="checkbox"/> Yes

Figure 7. RAST settings used to annotate the genomes.

<http://rast.nmpdr.org/>

After sequences were annotated, they were submitted to GenBank. It should be mentioned that the sequences currently on GenBank have been reannotated with NCBI's Prokaryotic Genome Annotation Pipeline (PGAP), a service available for GenBank submitters. NCBI's annotation pipeline uses several NCBI databases and is not available for public download (31).

III-c-ii Genome annotation comparison

To compare the annotations of the 12 previously sequenced VA clinical isolates, progressiveMauve was used to align the sequences. ProgressiveMauve is a free downloadable tool that creates positional homology alignments and excels in aligning rearranged genomes (32). To do this, the 12 previously sequenced VA isolates were downloaded from GenBank. The sequences can be found by searching NCBI's BioProject database for accession number PRJNA321467. Once the GenBank sequences were downloaded, they were input into progressiveMauve along with the GenBank sequence for reference strains Tohama I E4786 and Chinese reference strains (CS) C393 for alignment using default parameters. Tohama I and C393 were chosen because they are some of the strains that are used to derive the protein components used in aP vaccines (33).

III-d Structural similarity tree construction

III-d-i Permutations output file and data formatting.

To view the structural similarity of the sequences a permutations output file was first generated from the progressiveMauve alignment by clicking Tools → Export → Export permutations. Next, Maximum Likelihood for Gene Order Analysis (MLGO) was performed. To do this, the file was reformatted using the search and replace tool in Visual Studio Code to switch commas to spaces and add name labels. The original progressiveMauve file and the edited MLGO input format file can be accessed in the supplemental materials section.

III-d-ii Submission of job to MLGO and creation of tree using R

Next, the MLGO input format file was uploaded into the input data section of MLGO web server hosted at www.geneorder.org (34). The interference target was selected to be phylogenetic tree reconstruction. The progressiveMauve guide tree was then input under the small phylogeny problem only section.

To visualize the MLGO output tree, the .tree file was downloaded from the zip file. An R script, shown below, that utilizes the tidyverse CRAN package and Bioconductor ggtree package was then used to visualize the tree (35).

```
library(tidyverse)
library(ggplot2)
library(ggtree)
t<-read.tree("MLGOresults.tree")
plot(t)
```

Figure 8. Script that was used to create structural similarity tree

III-e MLST and virulence factor allele typing

III-e-i Overview of BIGSDB and PubMLST

To compare our sequences to other sequences MLST was performed using www.pubmlst.org (PubMLST) *Bordetella spp.* database (36). PubMLST is a site hosted by The Department of Zoology at the University of Oxford with funding from The Wellcome Trust (36). PubMLST is powered by the BIGSDB software platform, which is an open source, online-hosted system that links phenotype and sequence data from whole genomes to bacterial specimens (37). PubMLST's *Bordetella* database uses two linked databases. The Sequence definitions database contains allele sequences and MLST profile definitions and the Isolate database contains epidemiological information (36).

III-e-ii MLST and sequence typing

MLST and strain typing was performed based on seven housekeeping genes described previously (38). The seven housekeeping genes used in the *Bordetella* MLST are *adk*, *fumC*, *glyA*, *tyrB*, *icd*, *pepA*, and *pgm*. To perform MLST and strain typing, PubMLST was accessed and the following process was used: Databases → *Bordetella spp.* → Sequence and profile definitions → Sequence query → select MLST scheme → choose strain FASTA file → Submit. After the FASTA file was submitted, the screen returned the allele types for each of the seven housekeeping genes. The combination of these alleles was used to determine the sequence type (ST). The MLST profile and ST was recorded for each isolate.

III-e-iii Virulence factor allele typing

To understand how virulence factor genes may be evolving due vaccine selection pressure, several virulence factors commonly used in molecular typing were analyzed (39). The loci analyzed were *prn*, *ptxS1* (*ptxA*), *ptxP*, and *fim3* (*fimH*). These loci were chosen because they encode antigens commonly used in aP vaccines and may shed light onto selective pressures caused by the vaccine (40). To perform molecular typing, PubMLST was accessed and the following process was used: Databases → *Bordetella spp.* → Sequence and profile definitions → Sequence query → select All loci → choose strain FASTA file → Submit. Once the job was complete, loci for the specified genes were searched and the allele number for each gene was

recoded.

After the alleles were determined from the nucleotide fasta sequences, the genbank (.gbk) files were used to determine the protein variant of the loci. This was done because there are more nucleotide alleles than protein variants due to nucleotide mutations that do not change the protein sequence (41). To determine protein variants, the protein sequences of Prn, PtxS1, and Fim3 were downloaded from the VA .gbk files then compared to known sequences for each protein variant that were described in previous studies using ClustalOmega (42, 43). The information from this molecular typing analysis was then compared to similar studies done in the United States and globally.

IV Results and Discussion

IV-a ProgressiveMauve whole-genome alignment

The annotation and alignment of the genomes indicated that there was little variation in the genomic content of the Va isolates between the genomes. However, while all the genomes have similar genetic content, there is significant genomic rearrangement between the Va clinical isolates. ProgressiveMauve illustrates rearrangements with locally colinear blocks (LCBs) that represent regions in different genomes that occur in the same order and orientation (32). Each row in the alignment represents a strain and the LCBs are illustrated by blocks of the color in different genomes (32). The lines that connect the LCBs illustrate the relative position of an LCB in one genome to the same LCB in another genome (32). The progressiveMauve alignment of temporarily different Va clinical strains along with Tohama 1 and C393, as shown in figure 9, indicates significant rearrangement of the genomes shown by the pronounced crisscrossing of the LCB connecting lines. Recent studies using similar methods have shown that the majority of these rearrangements are caused by large inversions at the replication origin or terminus that are bordered by IS481 mobile elements (39).

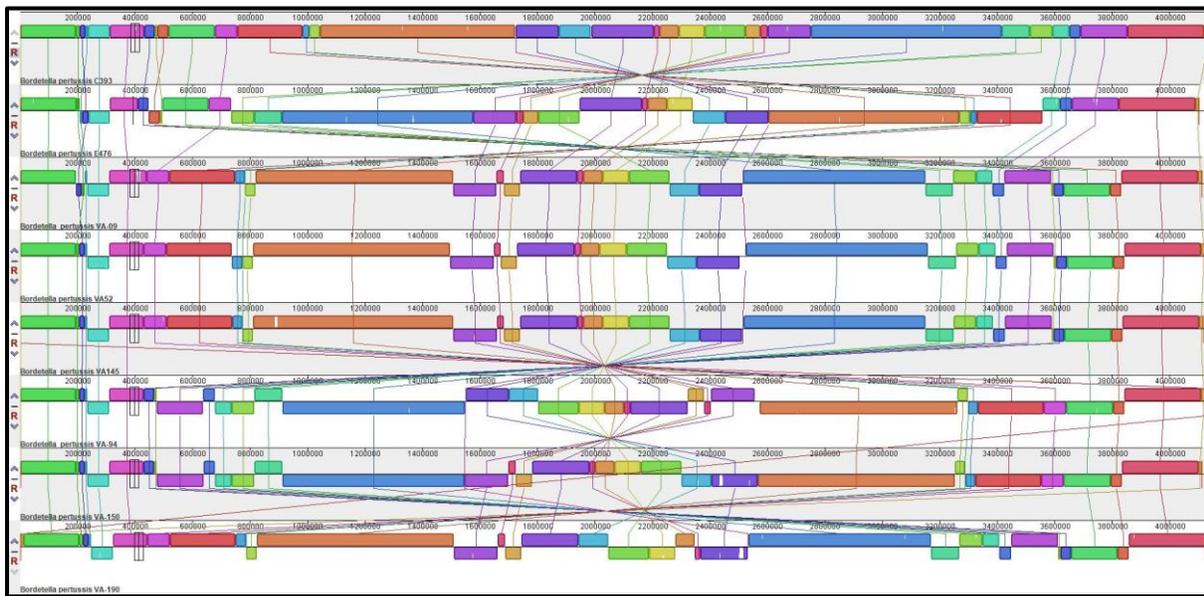


Figure 9. Figure showing progressiveMauve alignment 5 VA strains and vaccine reference strains based of local collinear blocks (LCBs). The vaccine references trains, C393 and Tohama 1, are the first two strains and VA-09, VA-52, VA-145, VA-94, VA-150, and VA-190 follow. The LCBs in the different genomes are indicated by different colors. The lines connecting the LCBs indicated the relative position of the LCBs in the different genomes. These connecting lines show large internal inversions within the strains.

IV-b Cladogram based on LCB order

The progressiveMauve permutations file which contains LCB orders for each genome was input into MLGO to create a tree file. This tree file was then downloaded, and an R script was used to create the cladogram that is shown below in figure 10 (35).

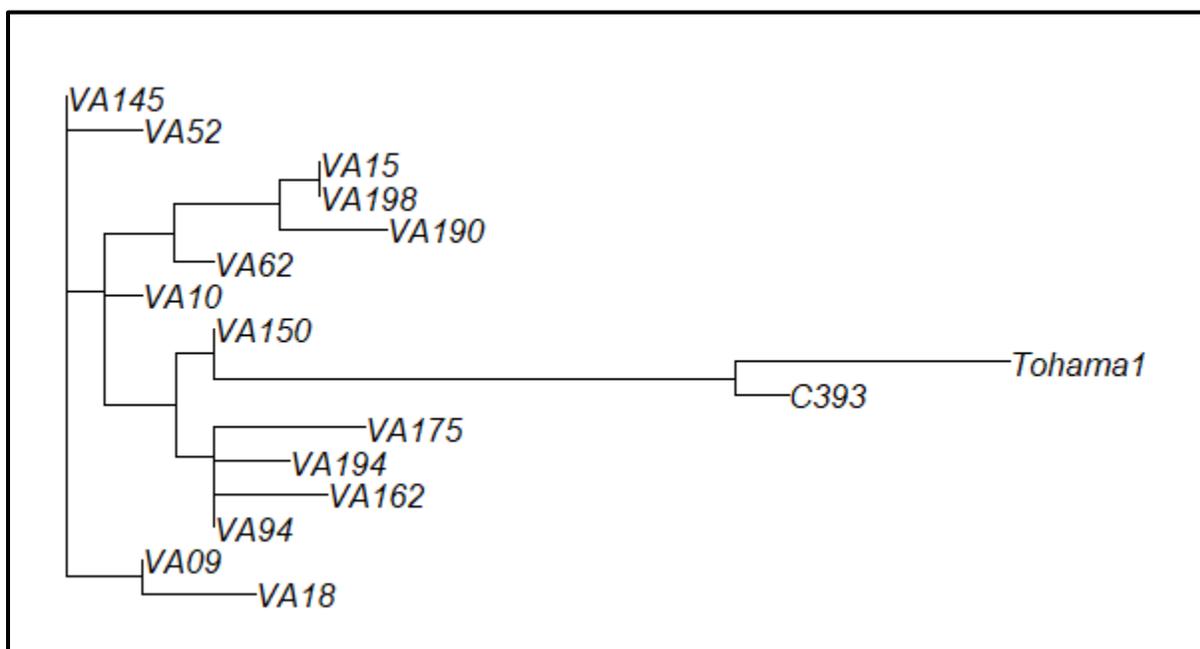


Figure 10. Cladogram based on LCB order of VA clinical Isolates and Tohama / C393 vaccine reference strains.

In figure 10 above, the length of the branches indicate the magnitude of rearrangements between the strains (44). The noticeably longer branch of Tohama 1 and C393 indicates that the Va clinical isolates have undergone significant rearrangements compared to vaccine reference strains (44). This indicates that the structural arrangement of Tohama 1 and C393, which are commonly used reference strains used to produce vaccines, may not be representative of circulating *B. pertussis* strains in Virginia (45).

IV-c Multi-Locus Sequence Typing (MLST)

MLST was performed using www.pubmlst.com according the allele sequences of seven housekeeping genes (*adhA*, *fumC*, *glyA*, *tyrB*, *icd*, *pepA* and *pgm*) that were described in Diavatopoulos et al 2005 (38). All the 14 Va strains in this this study except for VA-190 and VA-198 were determined to be sequence type (ST) 2.

This is consistent with current literature that suggest ST2 became the dominate sequence type in

the late 1990s as opposed to ST1, the sequence type of reference strains such as Tohama 1 (46). This change in ST is characterized by a change in the *tyrB* allele from 1 to 3 (47).

VA-190 and VA-198 were determined to ST83. This is a new strain type that was entered into the pubmlst database in February of 2018. This strain type profile is the same as ST2 except for a change in *icd* allele from 1 to 11. The ST1, ST2, and ST83 and the current number of isolates in the pubmlst database associated with each sequence type can be seen below in figure 11.

ST	adk	fumC	glyA	tyrB	icd	pepA	pgm	42 isolates
1	1	1	1	1	1	1	1	
ST	adk	fumC	glyA	tyrB	icd	pepA	pgm	905 isolates
2	1	1	1	3	1	1	1	
ST	adk	fumC	glyA	tyrB	icd	pepA	pgm	24 isolates
83	1	1	1	3	11	1	1	

Figure 11. Figure showing the allele numbers, and the current number of isolates in the pubmlst isolate database for ST1, ST2, and ST83

Some literature suggest the emergence of ST2 as the dominate sequence may be associated with the introduction of aP vaccines in the 1990s (48). Most sequences in this study were ST2, which are consistent with circulating strains mentioned in other studies (46). However, two of the strains were ST83, which is a new designation that has not been discussed in published literature. While there are 24 other isolates in PubMLST's database with ST 83, only one other isolate has a date associated with its isolation, 2010. This is only a few years prior to when the Va ST 83 strains were isolated, 2013.

IV-d Virulence factor allele typing

Allele and protein variants for pertactin (*prn*), pertussis toxin subunit 1 (*ptxS1*), pertussis toxin

promoter (*ptxP*), and fimbriae 3 (*fim3*) were analyzed. These genes were analyzed because they code for surface proteins, their products are often used as components in aP vaccines, and because they have shown divergence from vaccine reference strains (46, 49). Strain variation has been shown to alter vaccine efficacy in murine models (50). The allele types, derived from PubMLST, of the selected virulence factors can be seen in table 1.

Table 1. PubMLST allele designations and predicted protein variants. PubMLST nucleotide allele designations may not be the same as designations commonly used in *B. pertussis* studies. The highlighted strains are the strains sequenced in this study and the vaccine reference strains are bold. Predicted protein variants for Prn, PtxA, and FimH were found by aligning proteins with known reference alleles previously described (51)

Strain	<i>prn</i> ^a	Prn ^b	<i>ptxA</i> ^{c,d}	PtxA	<i>ptxP</i>	<i>fimH</i> ^{e,f}	FimH	Year	PFGE
VA-09	9	2	1	1	3	3	2	2001	vabpx002
VA-010	9	2	1	1	3	3	2	2001	vabpx001
VA-15	9	2	1	1	3	3	2	2001	vabpx001
VA-18	9	2	1	1	3	3	2	2001	vabpx002
VA-52	9	2	1	1	3	3	2	2005	vabpx002
VA-62	9	2	1	1	3	3	2	2005	vabpx008
VA-94	9	2	1	1	3	1	1	2008	vabpx007
VA-145	9	2	1	1	3	3	2	2010	vabpx002
VA-150	NULL	-	1	1	3	1	1	2010	vabpx023
VA-162	NULL	-	1	1	3	1	1	2011	vabpx007
VA-175	NULL	-	1	1	3	1	1	2012	vabpx018
VA-190	NULL	-	1	1	3	3	2	2013	vabpx045
VA-194	NULL	-	1	1	3	1	1	2012	vabpx008
VA-198	NULL	-	1	1	3	3	2	2013	vabpx030
Tohama 1	1	1	4	2	1	1	1	1954	CDC232
C393	1	1	4	2	1	1	1	1951	CDC052

^a NULL in this column means PubMLST did not return any values for *prn*

^b - in this column indicates functional pertactin was not expressed

^c *ptxA* is also designated *ptxS1*

^d alignment of Tohama 1 and C393 *ptxA* allele with a reference alleles showed PubMLST's *ptxA4* designation is the same sequence as *ptxA2* that has been characterized in literature (51).

^e *fimH* is also designated *fim3*

^f alignment of *fimH3* alleles with reference alleles showed that PubMLST's *fimH3* designation is the same sequence as *fimH2* that has been characterized in literature (51)

IV-d-i Pertactin

Pertactin is one of the most variable genes in the *B. pertussis* genome (52). Strains used to derive aP vaccines, such as Tohama 1, contain the *prn1* allele which encodes for the Prn1 antigenic variant (12). Today, Prn2 strains dominate most vaccinated populations worldwide (10). Prn deficient strains are becoming increasingly common among vaccinated populations (53). The ability of *B. pertussis* to infect without one if its presumed critical virulence factors suggests that the bacterium may be evolving to evade the immunity produced by the vaccine (33).

The PubMLST results indicate that eight of the strains are *prn9*. However, when the protein sequences of these strains were compared to known references with ClustalOmega, all *prn9* strains were determined to be Prn2. This may be because most variation of *prn* alleles occurs in repeat motifs that do not alter the amino acid sequence (53).

PubMLST indicated that six of the strains were *prn* deficient. A manual check of the .gbk sequences gave insight into potential reasons why these strains did not express Prn. A table showing the notes for pertactin in the .gbk files for the strains can be seen below in table 2.

Table 2. Table showing notes for pertactin in the .gbk files of pertactin deficient strains.

Notes were created during the PGAP and derived by automated computational analysis using gene prediction method: Protein Homology.

Strain	Start	Stoop	Note
VA150	1079795	1082542	internal stop;
VA162	3926910	3928184	Unknown disruption
	3928248	3929657	Unknown disruption
VA175	1079802	1081415	incomplete; partial in the middle of a contig; missing stop;
	1082462	1083598	incomplete; partial in the middle of a contig; missing start;
VA190	3322295	3325040	frameshifted;
VA194	1233042	1235549	incomplete; partial in the middle of a contig; missing start;
VA198	114646	117391	frameshifted;

The Prn analysis of our strains demonstrated that the Prn1 component used in most acellular vaccines is not representative of the Prn protein variant of strains circulating in Va. The Va strains in this study either exhibited Prn2 or were Prn deficient.

IV-d-ii Pertussis toxin subunit 1 (*ptxA*)

Pertussis toxin (PT) is a complex toxin made up of multiple subunits (one active subunit and five binding subunits) (11). Pertussis toxin subunit 1 (*ptxA*), also designated *ptxS1*, is the first gene in the PT operon and where the enzymatic activity of PT resides (12, 11). All 14 of the Va strains exhibited a *ptxA1* allele. This is consistent with other studies that claim around 90% of global circulating strains harbor a *ptxA1* allele (64.). This is in contrast to vaccine reference strain Tohama 1 harbors a *ptxA2* allele which encodes for the PtxA2 (55). PtxA1 is thought to have replaced PtxA2 as the predominant variant after the introduction of whole-cell whooping cough vaccines (10). The rapid replacement of PtxA2 with PtxA1 as the dominant allele indicates that the bacterium may be evolving to escape vaccine induced immunity and increase strain fitness (54).

IV-d-iii Pertussis toxin promoter (*ptxP*)

Unlike the other virulence factors studied in this analysis, *ptxP* has no protein product that is used to produce aP vaccines (56). However, *ptxP* regulates the expression of *ptxA* and certain *ptxP* alleles have been shown to significantly increase the expression of PT (51). Tohama 1 and other vaccine strains contain a *ptxP1* allele (12). All of the 14 Va strains exhibited a *ptxP3* allele. This is consistent with other studies that have found that since their detection in the 1990s, *ptxP3* strains have increased in frequency and now represent up to 90% of circulating strains in some areas (41). In one study, *PtxP3* was shown to increase expression of PT by a factor of 1.6 relative to *ptxP1* (51). This increased production of PT increases may give *ptxP3* strains a selective advantage over *ptxP1* strains which could explain the predominance of *ptxP3* strains (57). It has been hypothesized that the increased production of PT may increase the severity of the disease leading to higher hospital rates and subsequently a higher reported number of *ptxP3* strains (57).

IV-d-iv Fimbriae 3 (*fim3*).

Fimbriae serotype 3 (*fim3*), also *fimH*, is another gene that encodes Fim3 which is a component of many acellular vaccines(58, 59). *Fim3-1*, *fimH1*, and *fim3A* are all used in different studies to refer the same nucleotide sequence for a *fim* allele(39, 58). The PubMLST typing results for *fim3* (*fimH*) can be seen in table 1.

It is important to note that because of the way PubMLST assigns allele designations, the PubMLST allele type may not be the same designation as the allele type used in current literature. An example of this is seen in table 1. While PubMLST calls *fimH3* for VA-09, VA10, VA15, VA-15, VA-18, VA-52, VA-62, VA-145, VA-190, and VA-198, there nucleotide sequence is the same as *fim3-2* (*fim3B*) that is mentioned in literature (58). This is because in the literature there is a *fim3-1** (*fim3A**) designation that encodes for the same protein as *fim3-1* but has a silent polymorphism at that changes C to T at position 29 of the nucleotide sequence (58).

Tohama 1 contains a *fim3-1* allele. Our results are consistent with other studies that show a correlation of increased cases and the emergence of *fim3-2* from the early 2000s until around 2010 (12). However, also consistent with other studies, there seems to be an unexplained reemergence of the *fim3-1* allele around 2010 (12).

V. Conclusions and future work

Pertussis is the most poorly controlled vaccine-preventable disease, and the number of cases has been on the rise since the 1990s. This study investigated bacterial evolution as a possible reason for this resurgence. In this study, a total of fourteen VA clinical *B. pertussis* isolates were compared to each other and to vaccine reference strains Tohama 1 (E4786) and CS (C393). An analysis of gene clustering using the order and orientation of homologous sequence blocks showed that the vaccine strain and other strains with the *fimH1* group together.

Reference vaccine strains used to produce acellular vaccines are not representative of circulating strains. For this reason, several genes that encode for vaccine components were also analyzed. The analysis showed that Virginia *B. pertussis* strains have *prn*, *ptxA*, *ptxP*, and *fim3* allele types that are not only divergent from vaccine strain allele types but also consistent with the allele types mentioned in other studies on globally circulating strains.

While each strain had PFGE data associated with it, this data was not incorporated in this analysis due to troubles with interlaboratory comparison of PFGE results. Originally, I wanted to do an unbiased SNP search to see if the frequency of SNPs in virulence genes in relation to other functional categories. Factors that I believe limited my success in this were these: 1. Lack of funds for proprietary software (i.e. CLC genomics workbench), 2. inexperience with the Linux operating system, since many programs are in Linux. , and 3. Many programs are not designed to handle long-reads (i.e. Geneious).

There is a significant amount of future work to be done with these and other *B. pertussis* strains. One thing that may be useful is to develop a standardized pipeline for sequencing, annotating, and characterization of *B. pertussis* isolates. This will improve the ability to study strains from different temporal periods and geographical regions to better understand the evolution of the bacterium.

Whole-genome multi-locus sequence typing is one future research avenue that is being developed by the CDC. While this pipeline is still in preliminary phases, and there is no published work on it, the CDC is investigating a whole-genome multi-locus sequence typing scheme that looks at around 3600 targets instead the 5 that are currently investigated. This would provide a better understanding of evolution across the entire genome and not just a select few virulence factors.

Another future research avenue is studying potential new vaccine antigens. The evidence that most recent circulating strains are Prn deficient indicates that some of the current aP vaccine antigens may not be essential. This means that novel vaccine antigens should be explored to potentially improve the vaccine. However, one caveat with developing a new vaccine is that the pertussis vaccine is currently a component diphtheria, tetanus, and pertussis (DTaP) vaccine. This means that changing the pertussis protein components would require the development of a completely new vaccine which could be expensive.

VI. Acknowledgements and funding

I would like to thank my advisor Dr. Louise Temple for the significant investment of time and energy that she contributed to help me with this project. I would like to thank my collaborators at UVA, Dr. Josh Eby and Dr. Laura Gonyar, for the invaluable lab experience they provided me with. I would also like to thank my committee members, Dr. Amanda Biesecker and Dr. Steven Cresawn, I appreciate the feedback and revisions. I would also like to thank all members of Temple lab for the assistance and friendship.

A 4-VA grant titled "A Virginia collaborative effort to analyze genomes of recent whooping cough bacteria" awarded to Dr. Temple and Dr. Eby in 2016 provided funding for this research.

Bibliography

1. Sealey KL, Belcher T, Preston A. 2016. *Bordetella pertussis* epidemiology and evolution in the light of pertussis resurgence.
2. James Cherry PD. *Pertussis in the Preantibiotic and Prevaccine Era, with Emphasis on Adult.*
3. Bordet J, Gengou O. 1906. No Title. *Le Microbe la coqueluche* 20:48–68.
4. Cherry JD. Historical Review of Pertussis and the Classical Vaccine.
5. 2015. Weekly epidemiological record *Relevé épidémiologique hebdomadaire* 90:433–460.
6. CDC, Ncird. *Immunology and Vaccine-Preventable Diseases – Pink Book – Pertussis.*
7. Cherry JD. 2013. Pertussis: Challenges Today and for the Future. *PLoS Pathog* 9:9–12.
8. Edwards DJ, Holt KE. 2013. Beginner’s guide to comparative bacterial genome analysis using next-generation sequence data. *Microb Inform Exp* 3:2.
9. Nieves DJ, Heininger U. 2016. *Bordetella pertussis*, p. 311–339. *In* Emerging infections 10. American Society of Microbiology.
10. Dorji D, Mooi F, Yantorno O, Deora R, Graham RM, Mukkur TK. 2017. *Bordetella Pertussis* virulence factors in the continuing evolution of whooping cough vaccines for improved performance. *Med Microbiol Immunol* 0:1–24.
11. Carbonetti NH. 2010. Pertussis toxin and adenylate cyclase toxin: key virulence factors of *Bordetella pertussis* and cell biology tools. *Future Microbiol* 5:455–69.
12. Bowden KE, Williams MM, Cassiday PK, Milton A, Pawloski L, Harrison M, Martin SW, Meyer S, Qin X, Debolt C, Tasslimi A, Syed N, Sorrell R, Tran M, Hiatt B, Lucia M. 2014. Molecular Epidemiology of the Pertussis Epidemic in Washington State in 2012 *52:3549–3557.*
13. SEPKOWITZ S. 1992. The Artful Dodge on Pertussis Vaccine. *Arch Pediatr Adolesc Med* 146:788.
14. Cherry J, Harriman K. 2012. Why Do Vaccine-Preventable Disease Outbreaks Occur in the US? *Infect Dis Spec Ed* 53–57.

15. Fine PEM. Herd Immunity: History, Theory, Practice 15.
16. Anderson RM, May RM. 1985. Vaccination and herd immunity to infectious diseases. *Nature* 318:323–329.
17. Castagnini LA, Healy CM, Rench MA, Wootton SH, Munoz FM, Baker CJ. 2012. Impact of Maternal Postpartum Tetanus and Diphtheria Toxoids and Acellular Pertussis Immunization on Infant Pertussis Infection. *Clin Infect Dis* 54:78–84.
18. Healy CM, Rench MA, Castagnini LA, Baker CJ. 2009. Pertussis immunization in a high-risk postpartum population. *Vaccine* 27:5599–5602.
19. Weiss AA, Hewlett EL. 1986. Virulence Factors of *Bordetella Pertussis*. *Annu Rev Microbiol* 40:661–686.
20. Romano MJ, Weber MD, Weisse ME, Siu BL. 2004. Pertussis pneumonia, hypoxemia, hyperleukocytosis, and pulmonary hypertension: improvement in oxygenation after a double volume exchange transfusion. *Pediatrics* 114:e264-6.
21. De Berry BB, Lynch JE, Chung DH, Zwischenberger JB. 2005. Pertussis with severe pulmonary hypertension and leukocytosis treated with extracorporeal membrane oxygenation. *Pediatr Surg Int* 21:692–694.
22. Deeks S, De Serres G, Boulianne N, Duval B, Rochette L, Déry P, Halperin S. 1999. Failure of Physicians to Consider the Diagnosis of Pertussis in Children. *Clin Infect Dis* 28:840–846.
23. Sotomayor J, Weiner LB, McMillan JA. 1985. Inaccurate diagnosis in infants with pertussis. An eight-year experience. *Am J Dis Child* 139:724–7.
24. Girard DZ. 2012. Recommended or mandatory pertussis vaccination policy in developed countries: Does the choice matter? *Public Health* 126:129–134.
25. Producing Prevention: The Complex Development of Vaccines - Blog.
26. Pfizer. 2014. The Value of Vaccines in Disease Prevention.
27. Temple L, Eby J. 2015. A Virginia Collaborative Effort to Analyze Genoms of Recent Whooping Cough Bacteria.
28. Biosciences P. Revolutionize Genomics with SMRT® Sequencing Single Molecule, Real-Time Technology.
29. Eby JC, Turner L, Nguyen B, Kang J, Neville C, Temple L. 2016. Complete Genome

- Sequence of *Bordetella pertussis* Strain VA-190 Isolated from a Vaccinated 10-Year-Old Patient with Whooping Cough. *Genome Announc* 4:e00972-16.
30. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:206–214.
 31. NCBI. 2018. NCBI Prokaryotic Genome Annotation Pipeline.
 32. Darling AE, Mau B, Perna NT. 2010. progressiveMauve : Multiple Genome Alignment with Gene Gain , Loss and Rearrangement 5.
 33. Whooping RUS, Statewide C, Bowden KE, Weigand MR, Peng Y, Cassiday PK, Sammons S, Knipe K, Rowe LA, Loparev V, Sheth M, Weening K, Tondella ML, Williams M. Genome Structural Diversity among 31 *Bordetella pertussis* Isolates from Two Epidemics 1:1–15.
 34. Hu F, Lin Y, Tang J. 2014. MLGO: phylogeny reconstruction and ancestral inference from gene-order data. *BMC Bioinformatics* 15:354.
 35. Turner S. 2017. 4-VA JMU 2017 Biological Data Science Workshop.
 36. University of Oxford. MLST databases and software - PubMLST.org.
 37. Jolley KA, Maiden MC. 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595.
 38. Diavatopoulos DA, Cummings CA, Schouls LM, Brinig MM, Relman DA, Mooi FR. 2005. *Bordetella pertussis*, the causative agent of whooping cough, evolved from a distinct, human-associated lineage of *B. bronchiseptica*. *PLoS Pathog* 1:0373–0383.
 39. Weigand MR, Peng Y, Loparev V, Batra D, Bowden KE, Burroughs M, Cassiday PK, Davis JK, Johnson T, Juieng P, Knipe K, Mathis MH, Pruitt AM, Rowe L, Sheth M, Tondella ML, Williams MM. 2017. The History of *Bordetella pertussis* Genome Evolution Includes Structural Rearrangement. *J Bacteriol* JB.00806-16.
 40. 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.
 41. Mooi FR, Van Der Maas NAT, De Melker HE. 2014. Pertussis resurgence: Waning

- immunity and pathogen adaptation - Two sides of the same coin. *Epidemiol Infect* 142:685–694.
42. Jung S-O, Moon YM, Kim S-H, Sung HY, Kwon S-J, Kang YH, Yu JY. 2011. Multilocus Sequence Analysis of Housekeeping Genes and Antigenic Determinant Genes in *Bordetella pertussis* Strains Isolated in Korea. *Osong public Heal Res Perspect* 2:115–26.
 43. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
 44. Rambaut A. 2013. How to read a phylogenetic tree | epidemic.
 45. Caro V, Bouchez V, Guiso N. 2008. Is the Sequenced *Bordetella pertussis* strain Tohama I representative of the species? *J Clin Microbiol* 46:2125–8.
 46. Kim S-H, Lee J, Sung HY, Yu JY, Kim SH, Park MS, Jung S-O. 2014. Recent trends of antigenic variation in *Bordetella pertussis* isolates in Korea. *J Korean Med Sci* 29:328–33.
 47. Yang Y, Yao K, Ma X, Shi W, Yuan L, Yang Y. 2015. Variation in *bordetella pertussis* susceptibility to erythromycin and virulence-related genotype changes in China (1970–2014). *PLoS One* 10:1–12.
 48. Yang Y, Yao K, Ma X, Shi W, Yuan L, Yang Y. 2015. Variation in *bordetella pertussis* susceptibility to erythromycin and virulence-related genotype changes in China (1970–2014). *PLoS One* 10:1–13.
 49. Mooi FR, van Loo IHM, van Gent M, He Q, Bart MJ, Heuvelman KJ, de Greeff SC, Diavatopoulos D, Teunis P, Nagelkerke N, Mertsola J. 2009. *Bordetella pertussis* Strains with Increased Toxin Production Associated with Pertussis Resurgence. *Emerg Infect Dis* 15:1206–1213.
 50. Watanabe M, Nagai M. 2002. Effect of Acellular Pertussis Vaccine Against Various Strains of *Bordetella Pertussis* in a Murine Model of Respiratory Infection. *J Heal Sci* 48:560–564.
 51. Marchand-Austin A, Tsang RSW, Guthrie JL, Ma JH, Lim G, Crowcroft NS, Deeks SL, Farrell DJ, Jamieson FB. 2017. Short-Read Whole Genome Sequencing for Laboratory-Based Surveillance of *Bordetella pertussis*. *J Clin Microbiol* JCM.02436-16.

52. Mooi FR. 2010. *Bordetella pertussis* and vaccination: The persistence of a genetically monomorphic pathogen. *Infect Genet Evol* 10:36–49.
53. Preston A. 2016. The role of *B. pertussis* vaccine antigen gene variants in pertussis resurgence and possible consequences for vaccine development. *Hum Vaccin Immunother* 12:1274–1276.
54. Bart M, Harris S, Advanid A, Arakawae Y, Botterof D, V B, Cassiday P, Chiangj C, Dalbyk T, Fryl N, Gaillard M, van Genta M, Guiso N, Hallanderd H, Harvill E, He Q, Van Der Heide H, Heuvelman K, Hozbor D, Kamachi K, Yao S, Zhang S, Parkhill J, Mooi R. 2014. Global Population Structure and Evolution of *Bordetella pertussis* and Their Relationship with Vaccination. *MBio* 5:e01074-14.
55. Advani A, Hallander HO, Bondeson K, König V, Riffelmann M, Vahrenholz C, Guiso N, Caro V, Njamkepo E, He Q, Mertsola J, Mooi FR, Amersfoorth SCM Van, Schouls LM, Heide HGJ Van Der, van Amersfoorth SCM, van der Heide HGJ, Wirsing von König C-H. 2005. Analysis of *Bordetella pertussis* Populations in European Countries with Different Vaccination Policies Analysis of *Bordetella pertussis* Populations in European Countries with Different Vaccination Policies. *J Clin Microbiol* 43:2837–2843.
56. Marchand-austin A. 2015. *Bordetella pertussis* : Whole Genome Sequence Analysis and its Application to Public Health Laboratory Surveillance by.
57. Lam C, Octavia S, Bahrame Z, Sintchenko V, Gilbert GL, Lan R. 2012. Selection and emergence of pertussis toxin promoter ptxP3 allele in the evolution of *Bordetella pertussis*. *Infect Genet Evol* 12:492–495.
58. Gorringe AR, Vaughan TE. 2014. *Bordetella pertussis* fimbriae (Fim): Relevance for vaccines. *Expert Rev Vaccines* 13:1205–1214.
59. van Gent M, Heuvelman CJ, van der Heide HG, Hallander HO, Advani A, Guiso N, Wirsing von König CH, Vestrheim DF, Dalby T, Fry NK, Pierard D, Detemmerman L, Zavadilova J, Fabianova K, Logan C, Habington A, Byrne M, Lutyńska A, Mosiej E, Pelaz C, Gröndahl-Yli-Hannuksela K, Barkoff AM, Mertsola J, Economopoulou A, He Q, Mooi FR. 2015. Analysis of *Bordetella pertussis* clinical isolates circulating in European countries during the period 1998–2012. *Eur J Clin Microbiol Infect Dis* 34:821–830.