1   # Genomic evaluation of *Bordetella* spp. originating from Australia

2   Winkie Fong[1,3,4], Verlaine Timms[1,3], Eby Sim[1,2], and Vitali Sintchenko[1,2,3,4]

3

4   [1]Centre for Infectious Diseases and Microbiology – Public Health, Westmead Hospital,
5   Westmead, NSW, Australia
6   [2]Centre for Infectious Diseases and Microbiology Laboratory Services, Westmead Hospital,
7   Westmead, NSW, Australia
8   [3]Westmead Clinical School, The University of Sydney, Westmead, NSW, Australia
9   [4]Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney,
10   Camperdown, NSW, Australia

11

12   Corresponding author: Winkie Fong – wfon4473@sydney.edu.au

13

14   **Abstract**

15   *Bordetella pertussis* is the primary causative agent of pertussis, a highly infectious

16   respiratory disease associated with prolonged coughing episodes. Pertussis infections are

17   typically mild in adults, however in neonates, infections can be fatal. Despite successful

18   vaccine uptake, the disease is re-emerging across the globe, therefore it is critical to

19   determine the mechanism by which *B. pertussis* is escaping vaccination control. Studies

20   have suggested that significant changes have occurred in *B. pertussis* genomes in response

21   to whole cell and acellular vaccines. Continued molecular monitoring is therefore crucial for

22   public health surveillance.

23

24   High-resolution molecular surveillance of *B. pertussis* can be achieved through the

25   sequencing of the whole genome. In public health laboratories, whole genome sequencing

26   is primarily performed by short-read sequencing technologies as they are most cost-

27   effective. However short read sequencing does not resolve the extensive genomic

28   rearrangement evident in *Bordetella* genomes. This is because repeat regions present in

29   *Bordetella* genomes are collapsed by downstream analysis. For example, the *B. pertussis*

30   genome contains more than 200 copies of the IS*481* insertion element, hence assemblies

31   generally consist of >200 contigs. Advancements in long-read technologies however

32   increase the potential to circularise and close genomes by bridging the locations of the

33   IS*481* insertion element.

34

35   In this study, we aimed to contextualise the *Bordetella* spp. circulating in NSW, Australia and

36   assess their relationship with global isolates utilising core genome, SNP and structural

37   clustering analysis using long read technology. We report five closed genomes of *Bordetella*

38   spp. isolated from Australian patients. Two of the three *B. pertussis* closed isolates, were

39   unique with their own genomic structure, while the other structurally clustered with global

40   isolates. We found that Australian *B. holmesii* and *B. parapertussis* strains cluster with global

41   isolates and do not appear to be unique to Australia. Australian draft *B. holmesii* SNP

42   analysis showed that between 1999 and 2007, isolates were relatively similar, however

43   post-2012, isolates were distinct from each other. The closed isolates can also be used as

44   high-quality reference sequences for both surveillance and other investigations into

45   pertussis spread.

46

47   **Keywords**

48   Bordetella pertussis, Bordetella holmesii, Bordetella parapertussis, Nanopore, Illumina

49

50   **Background**

51   Pertussis, or whooping cough, caused by *Bordetella pertussis* (and occasionally *Bordetella*

52   *parapertussis* and *Bordetella holmesii*) is a severe respiratory infection characterized by

53   prolonged coughing episodes[1]. In the pre-vaccination era, pertussis was the foremost cause

54   of infantile death due to an infectious disease in the first year of life[1]. Recently however,

55   pertussis has been re-emerging over the past few decades and this is despite high

56   vaccination coverage rates. The causes of this resurgence are possibly related to vaccine

57   escape, waning immunity or poor vaccine compliance. Further, with the rise of molecular

58   based diagnostic tests it is now clear that other *Bordetella* species such as *Bordetella*

59   *parapertussis* and *Bordetella holmesii* can also cause a mild form of pertussis-like illness. In

60   Australia, *B. holmesii* can circulate at a prevalence of between 6.5 - 16.8%[2] and is

61   indistinguishable from *B. pertussis* when using the common PCR diagnostic target of *IS481*[3].

62

63   Another possible factor contributing to the re-emergence of *B. pertussis* is a change in the

64   antigenic structure of circulating strains. Current acellular vaccines vary between countries

65   and contain 2-5 immunogenic antigens produced by *B. pertussis* – Pertussis toxin (PT),

66   Filamentous haemagglutinin (FHA), Pertactin (PRN), Fimbrae (FIM) and the adenylate

67   cyclase toxin (ACT). This has potentially led to the natural selection of circulating strains with

68   antigenic variants not present in the vaccine. Changes in antigenic variation have been

69   documented independently in several countries, including Australia, such as the emergence

70   of PRN-deficient[4, 5], FHA-deficient[6] or PT-deficient[7] strains of *B. pertussis* . It also appears

71   that over recent years, *B. pertussis* strains have diverged from historical clones to produce a

72   novel allele for the PT promoter *ptxP*[8]. Prior to the introduction of the vaccine, circulating

73   strains encoded *ptxP1,* however, current predominating strains contain *ptxP3*. Furthermore,

74   these recent *ptxP3* strains also produce 1.5 times more PT than their *ptxP1* counterparts[8].

75

76   Molecular surveillance of respiratory pathogens such as *B. pertussis* is paramount for

77   effective disease control[9]. Molecular typing of *B. pertussis* has been transformed with the

78   advent of higher-resolution genome sequencing methods such as whole genome

79   sequencing (WGS). Available isolates have allowed the development of single nucleotide

80   polymorphism (SNP) typing that is WGS-based. *B. pertussis* strains can be classified into SNP

81   clusters (I-VI)[10]. SNP clusters can be defined by specific SNP changes in genes, the primary

82   SNP change being in *ptxP* (G > A mutation in the intergenic region of *B. pertussis* Tohama I

83   BP3782 - designated *ptxP3*) and thus the sole definer of SNP Cluster I[10]. In Australia, SNP

84   Cluster I can be further divided into 5 epidemic lineages (EL1-EL5)[11].

85

86   WGS can provide a wealth of information that can be utilised for SNP-based typing or

87   observations of the vaccine antigen sequences using *de novo* assembly methods. However,

88   short-read based *de novo* assembly algorithms conglomerate repeat regions of the genome

89   resulting in basic and limited understandings of genomic structure and complexity[12]. This is

90   problematic in the assembly of the *B. pertussis* genome since it contains more than 200

91   copies of the IS*481* insertion element[13]. As a result, short-read assemblies are highly

92   fragmented, (often >200 contigs) which can lead to information being missed. Previous

93   research has also determined that the *B. pertussis* genome is subject to large genomic

94   inversions and rearrangements, with most mediated by *IS481*[14]. The type and positions of

95   the rearrangements are grouped by *B. pertussis* phylogenetic clusters[14]. However, these

96   structural rearrangements are not unique to *B. pertussis*, in fact, *B. parapertussis* and *B.*

97   *holmesii* also carry these characteristics[15].

98

99   Generating high-quality reference sequences that represent locally circulating strains can

100   more accurately identify variable regions of the genome, virulence and antibiotic resistance

101   markers and greatly support the accuracy of phylogenetic analysis[16, 17]. Advancements in

102   long-read technologies like Nanopore and PacBio sequencing have increased the potential

103   to circularise and close genomes by bridging the locations of the *IS481* insertion element[18-

104   20]. Further, as these genomic rearrangements have only been investigated in strains from

105   the northern hemisphere, primarily the US, the genomic rearrangements prevalent in

106   strains circulating in Australia and their significance for molecular surveillance is currently

107   unknown.

108

109   Therefore, this study sought to investigate the mechanisms of the potential evolution of

110   *Bordetella* spp. in Australia using a combination of long and short read technologies to

111   contextualise genomic rearrangements within previously defined SNP lineages. The study

112   present five closed *Bordetella* spp. genomes from the southern hemisphere that can be

113   used as references in future studies investigating *Bordetella* spp. strains circulating in

114   Australia.

115

116

117   **Results**

118   **Sequencing Statistics**

119   Five isolates were sequenced using both Nanopore long read sequencing and Illumina short

120   reads to close and polish the genome. Statistics of the sequencing results are presented in

121   Table 1.

122

123

124

125

126   Table 1: Summary statistics of closed *Bordetella* spp. genomes sequenced

Sample	Total Reads	Mean Read Length	Read Length N50	Genome Length	GC content (%)	Count of CDS	Count of tRNAs	Count of rRNA
CIDM-BH3	467,478	4,731.3	9,416	3,696,928	62.69	3,683	56	9
CIDM-BP2	501,426	6,830.1	14,169	4,109,518	67.70	3,971	64	9
CIDM-BP3	622,884	8,636.8	17,331	4,093,865	67.71	3,954	64	9
CIDM-BP5†	41,276	8,813	16,164	4,104,145	67.71	3,943	64	9
CIDM-BPP2	138,446	6,830.1	20,730	4,775,256	68.10	4,573	65	9

127   †Sequencing was only ran for 24 hours

128

129   In addition, we further sequenced seven *B. holmesii* isolates using short read technology

130   and sequencing and assembly statistics are listed in Table 2.

131

132   Table 2: Summary of the Illumina sequencing and assembly statistics of all *B. holmesii*
133   isolates.

| Sample	Total Reads	N50	Contigs (>1000 bp)	Genome Coverage (%)	GC content (%)	Count of CDS
CIDM-BH1	3,614,054	29,344	189	94.55	62.76	3,396
CIDM-BH2	3,432,966	29,426	188	93.87	62.79	3,557
CIDM-BH4	2,588,521	26,580	197	93.87	62.78	3,383
CIDM-BH6	2,542,224	27,377	197	93.91	62.78	3,386
CIDM-BH7	2,564,461	25,733	202	93.89	62.78	3,385
CIDM-BH8	2,939,553	28,704	193	93.89	62.78	3,382

134

135

136   **Sequenced *Bordetella pertussis* isolates exhibited IS*481* mediated genomic**

137   **rearrangements**

138   Three strains of *B. pertussis* were sequenced with long-read sequencing, and closure of each

139   genome was successful. *B. pertussis* strain CIDM-BP3 and CIDM-BP5 carried intact genes of

140   *prn* with CIDM-BP5 demonstrated to express PRN in previous work[21]. CIDM-BP2, however,

141   contained an IS*481* insertion at position 1598 of *prn*. *B. pertussis* strain CIDM-BP3 was a

142   *ptxP1* isolate, previously described as belonging to the non-Cluster I SNP cluster with a SNP

143   profile of 18 (SP18)[21]. Sequence alignments using the closed genome of CIDM-BP3

144   corroborated with the previous data[21]. Both *B. pertussis* strains CIDM-BP2 and CIDM-BP5

145   harboured the *ptx3* promoter and by definition[10] were both part of SNP Cluster I. Further

146   pairwise alignment analysis by BLASTn of CIDM-BP2 and CIDM-BP5, against key SNP

147   positions, classified them both as part of EL1, designated SP12 and SP13, respectively, due

148   to the SNP mutations in *gor, sphB3,* and BP3546.

149

150   The 23S ribosomal RNA was also investigated via *in silico* analysis and no mutations were

151   discovered in any *B. pertussis* strains. This indicates that these strains are macrolide

152   susceptible and is in line with phenotypical results for resistance against macrolides

153   (Supplementary Material). A selection of global closed *B. pertussis* isolates (n=192) (Figure

154   1) and previously described[14, 15] genomes used for structural clustering (n=472) (Figure S1)

155   were utilised for core phylogenomic and genomic rearrangement analysis respectively.

156   Results of both analyses revealed CIDM-BP2 aligned with more recent U.S.A and UK *ptxP3*

157   isolates and like CIDM-BP2, the international strains contained an IS*481* inserted at position

158   1598 of the PRN gene. In addition, CIDM-BP2 shared the same genomic rearrangements

159   previously observed in J247, I755, I228, I469, and I472 from Cluster-BP-12 (Figure 2).

160   Meanwhile CIDM-BP3 is closely related to *ptxP1* isolates from India (2017) and the U.K

161   (1967) with a fully intact *prn* gene. CIDM-BP3 was observed to be more similar to strains in

162   Cluster-BP-19 from the U.K (B226 and B228) rather than the United States (Figure S1). In

163   terms of genomic structural organisation, CIDM-BP3 possessed the same orthologous locally

164   collinear blocks (LCB) as H866, B226 and B228 (Figure 3) but differed from these isolates due

165   to IS*481* mediated inversions and hence designated as a singleton. Similarly, CIDM-BP5 was

166   also designated as a singleton due to it having a genomic organisation that was not

167   previously documented. In addition, inversions resulted in an insertion of an *IS481* element

168   downstream of *lysR* family transcriptional regulators in *ptxP3* isolates on two occasions,

169   once in CIDM-BP2 and once in CIDM-BP5 (Figure 3 and Supplementary material 2).

170
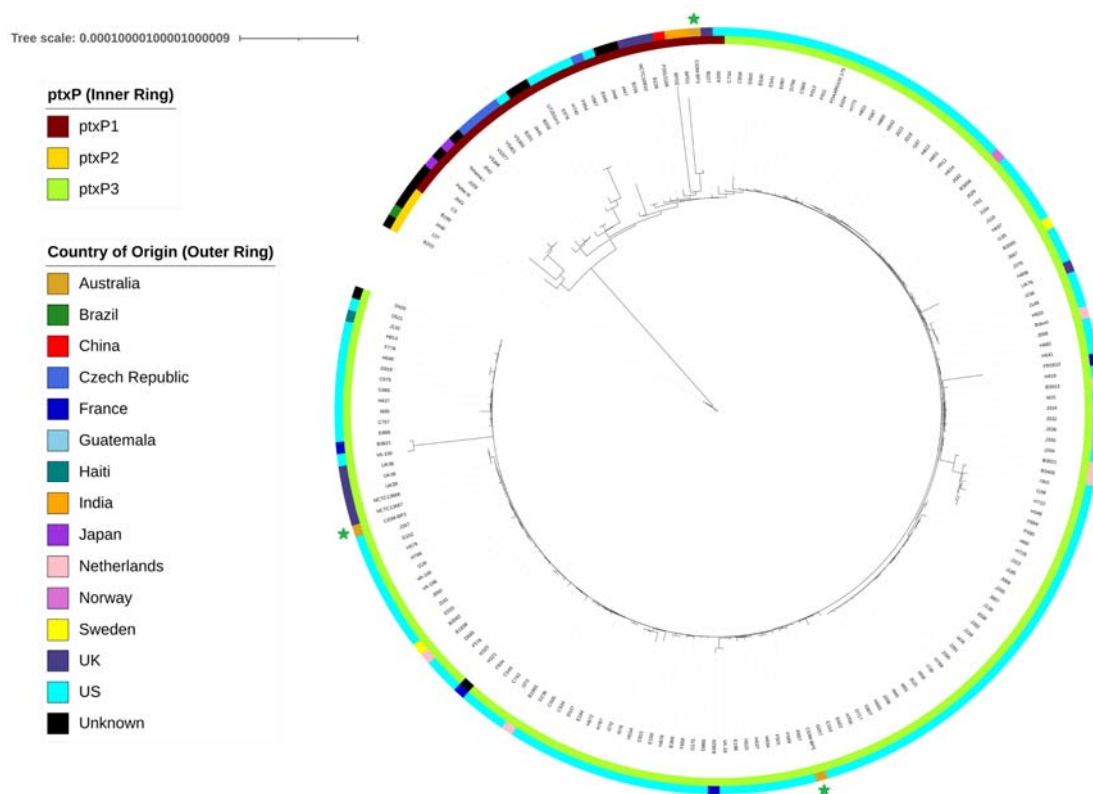
171
172   Figure 1: Core gene phylogeny of global closed *B. pertussis* isolates (n=192) based on 3,383
173   core genes using Roary. The tree shows that the *B. pertussis* isolates selected for long-read

174   sequencing in this study (green star) represent different strains that circulate globally. The

175   inner ring denotes *ptxP* alleles and the outer ring indicates country of origin.

176
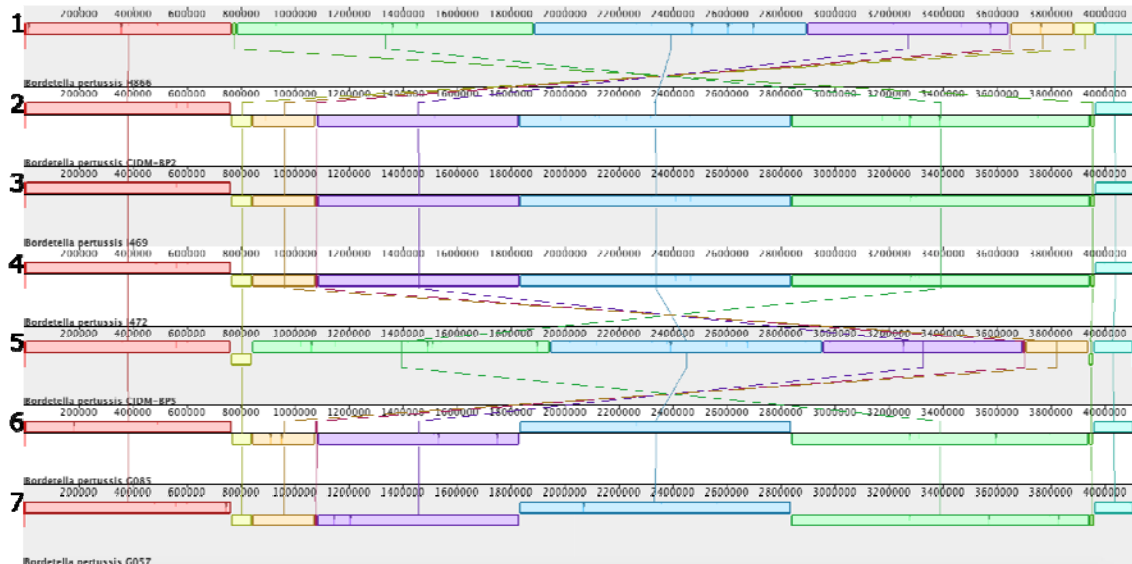
177

178   Figure 2: Genome structural variability comparisons of closed *B. pertussis ptxP3* genomes.

179   Australian strains CIDM-BP2 and CIDM-BP5 are in position 2 and position 5 respectively, and

180   are shown alongside their closest relatives by core gene phylogeny, I469 and I472 for CIDM-

181   BP2, and G085 and G057 for CIDM-BP5 (Figure 1). H866 was used to allow a comparison

182   with a previous study[14]. Image generated by progressiveMauve[22].
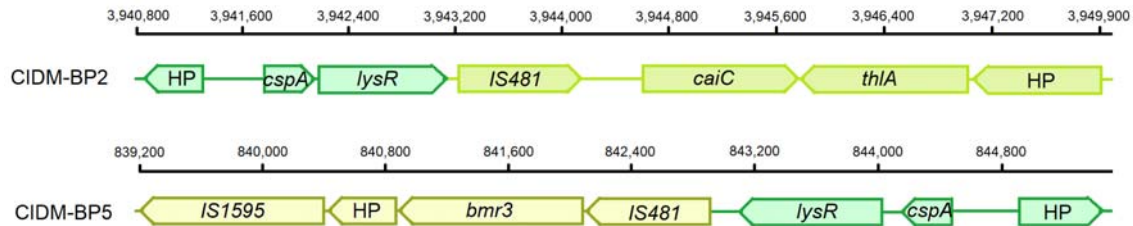
183
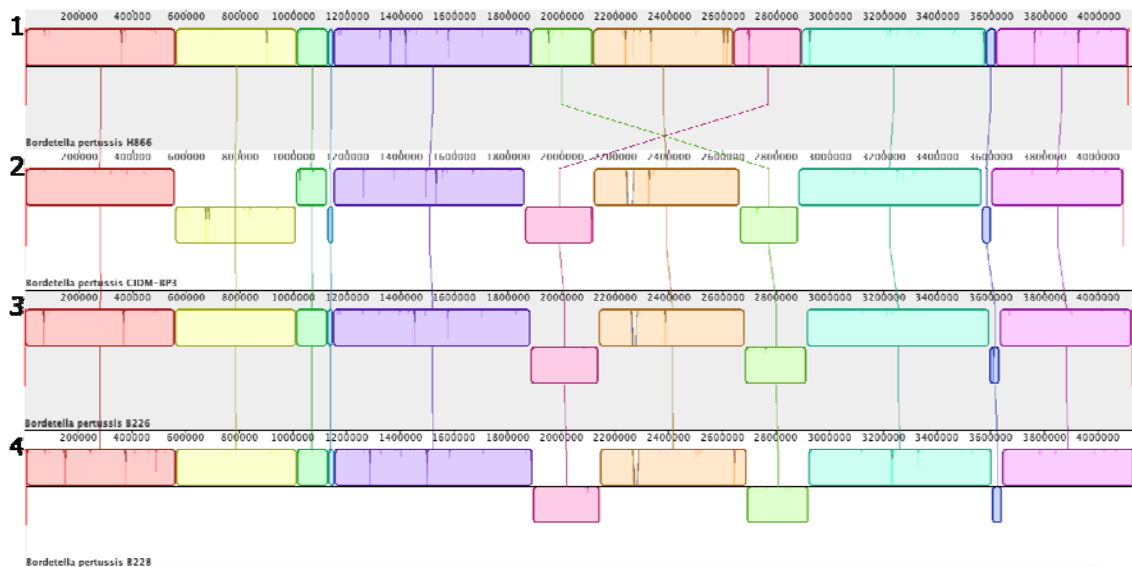
184

185   *Figure 3:* Structural rearrangement comparisons of closed *B. pertussis ptxP1* genomes, with

186   *B. pertussis* H866 as the anchoring genome The singleton CIDM-BP3 (position 2) is featured

187   in this figure comparing it to neighbouring isolates (B226 and B228) from the core gene
188   phylogeny. Image generated by progressiveMauve[22].
189

190

191
192   Figure 4: Schematic diagram of the two selected genomic rearrangement sites of CIDM-BP2
193   and CIDM-BP5. Represented here is the *IS481* element and three upstream and downstream
194   flanking genes, the colours are derived from the LCB blocks in Figure 2. In both examples the
195   *IS481* element has inserted downstream of *lysR*, as represented by the colours of the *IS481*.
196   The flanking gene names are as follows – hypothetical protein (HP), cold shock protein
197   (*cspA*), crotonobetaine/carnitine--CoA ligase (*caiC*), acetyl-CoA acetyltransferase (*thlA*) and
198   Multidrug resistance protein 3 (*bmr3*).
199

200

201   **CIDM-BPP2 possessed a genomic structure that was previously observed**

202   Comparison of CIDM-BPP2 to available RefSeq sequences (n=79) demonstrated that CIDM-

203   BPP2 clustered with B160 (Figure 5). Singleton B160 (and CIDM-BP2) differs from Cluster-PP-

204   02 by a 1,608 bp deletion in an Invasin gene (GenBank: AP019378.1). Therefore, CIDM-BPP2

205   forms a new cluster with B160. A loss of tandem repeats within the Invasin gene in Cluster-

206   PP-02, resulted in the truncation of gene length, however, appears to not disrupt gene

207   function. Examination of virulence markers of the *B. parapertussis* isolate revealed the

208   presence of two genes that were part of a type VI secretion system (T6SS), homologous to

209   *Pseudomonas aeruginosa* TssB1/HsiB1/VipA and TssC1/HsiC1/VipB T6SS (78% homology).

210   Further examination showed unlike the *B. parapertussis* Bpp5 reference genome (Genbank:

211   NC_018828.1) which contains only 7 of the 13 T6SS core components, CIDM-BPP2 and *B.*

212   *parapertussis* 12822 (Genbank: NC_002928.3) contains all 13 core components, and

213   designated T6SS Type I subtype i3 (Supplementary Table 1)

214

215

216
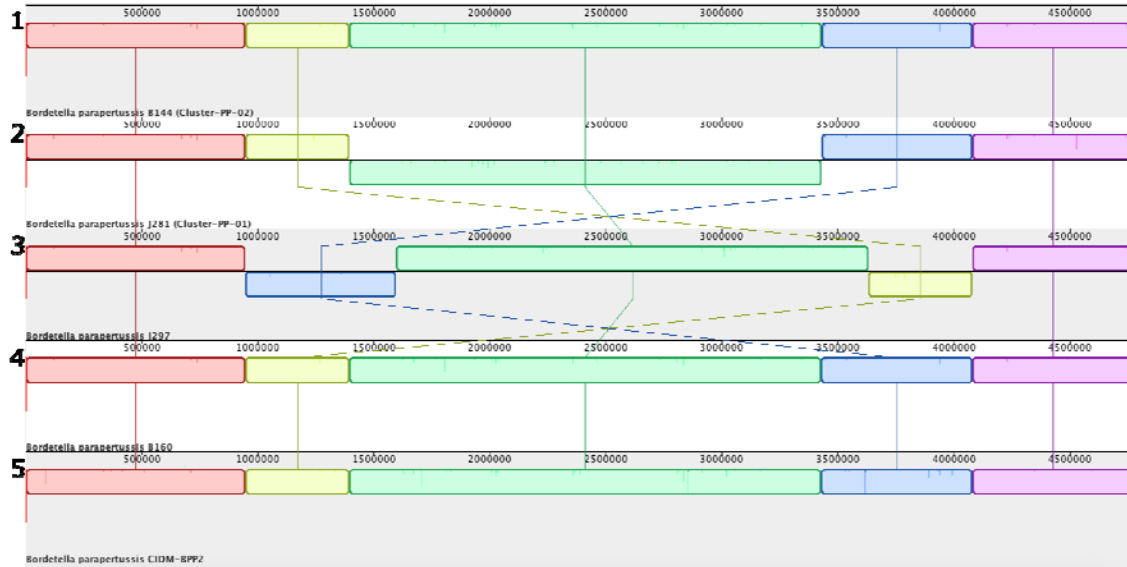
Figure 5: Structural rearrangement comparisons of all *B. parapertussis* structural clusters
referenced by a representative isolate selected from Cluster-PP-02 (position 1). J297
(singleton) is in position 3 as featured in a previous study and B160 (singleton) is in position
4 as it was the closest strain to CIDM-BPP2 (position 5) in the core phylogenetic analysis
(Figure S2). This figure shows that Cluster-PP-02 and B160 contain the same LCB
orientations, however, in depth analysis reveals a 1.6kb deletion in Cluster-PP-02 isolates
(genome coordinates 4,637,269 to 4,642,986). Image generated by progressiveMauve [22]

Global Comparison of *B. holmesii*

Phylogenetic analysis of all available *B. holmesii* isolates on RefSeq (n=84), SRA (n=10) and local (n=8), showed that CIDM-BH3 groups with Cluster-BH-01 as LCBs matched the genomic structure of the selected representative H869 (Figure 6). Like *B. pertussis* the structural rearrangements were mediated by *IS*481.

Further core genome analysis also disclosed that certain clusters display a similar absence of genes. A sub-cluster in Cluster-BH-01 is highly distinguished from other closed isolates due to a group of 53 missing genes. Half of the missing genes are IS elements, the remaining are listed in Supplementary Material 2. Whole genome screening of virulence markers using Abricate to interrogate VFDB, revealed that *B. holmesii* contains seven of the 10 homologs of the *B. pertussis bpl* locus, ranging from 76.39- 83.04% identity to 91.81- 98.86% coverage. *B. holmesii* lacks genes *bplG* (sugar transferase), *bplH* (glycosyl transferase) and *bplI* (lipopolysaccharide biosynthesis protein), which are genes involved in band-A lipopolysaccharide biosynthesis in *B. pertussis*[23]. Overall, *B. holmesii* lacks 87% of the virulence markers of *B. pertussis* as identified by VFDB and confirmed by BLASTn.

In total, there were 468 unique core SNP variants across Australian *B. holmesii* isolates (Supplementary Material 2). CIDM-BH1, CIDM-BH4, CIDM-BH6 and CIDM-BH7 formed a cluster (Figure 7) with 0 – 34 core SNPs between them. CIDM-BH1 and CIDM-BH4 differed from each other by 1 SNP, which is consistent as they were isolated from the same patient.
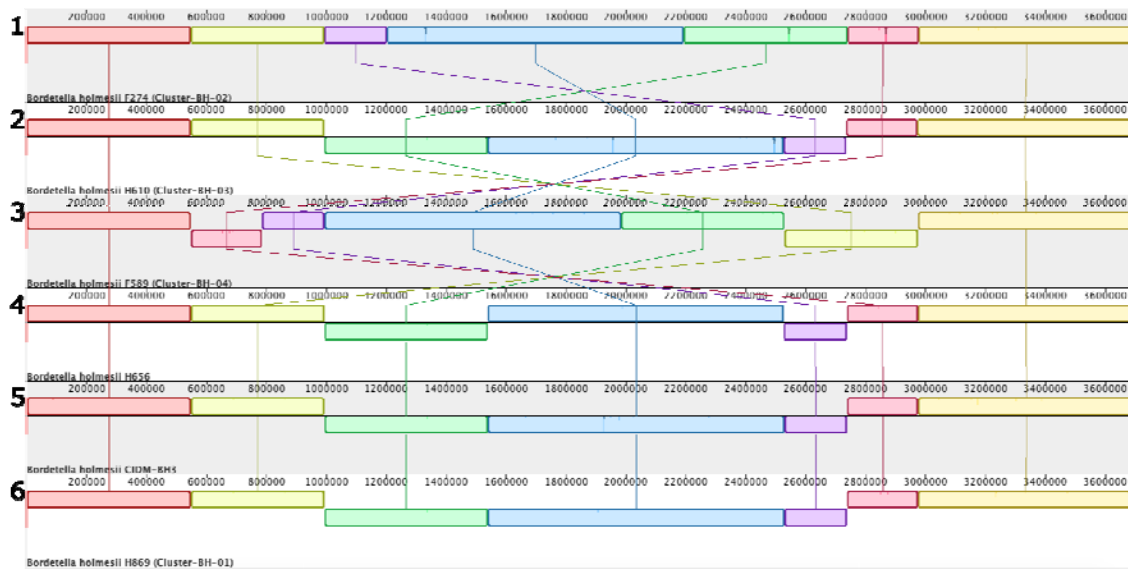

Figure 6: Genome structural rearrangements of *B. holmesii*. All structural clusters are featured, anchored by a representative isolate from Cluster-BH-02 (F274). CIDM-BH3 is featured in position 5, alongside its closest isolates H656 and H869 from Cluster-BH-01. Image generated by progressiveMauve[22].
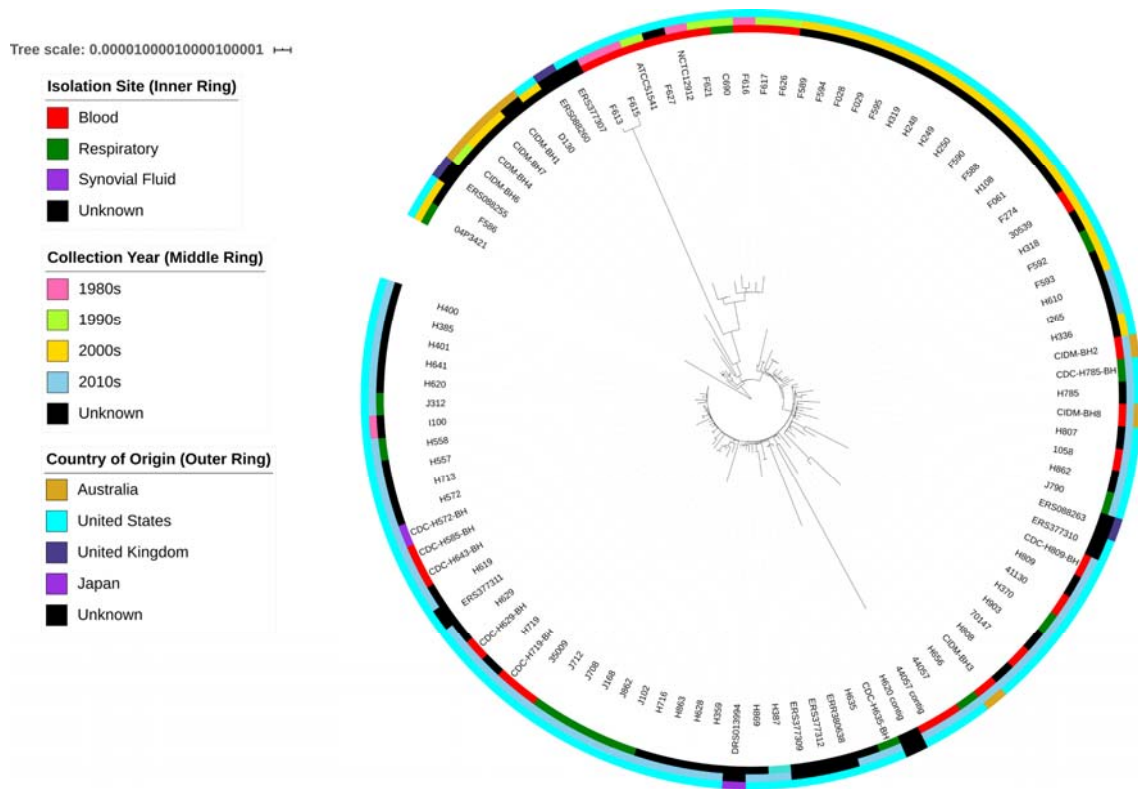
Figure 7: Core phylogenetic tree of all *B. holmesii* (n=101) used in the present study from NCBI SRA and RefSeq based on 2,747 core genes. Metadata is displayed in concentric circles with site of isolation (inner ring), year of collection (middle ring) and country of origin (outer ring) where available. The tree shows that Australian isolates pre-2007 are clustered together, while more recent isolates post-2012 are spread out across the globe with no specific grouping.

Discussion

Evidence of changes in the molecular epidemiology of *B. pertussis* have been prevalent across the globe[24]. Further research into identifying forces driving this change are critical to prevent and control the re-emergence of *B. pertussis.* This study aimed to contextualise Australian *Bordetella* spp. globally. By utilising both long and short-read sequencing technologies this study generated high quality reference sequences of circulating strains of *Bordetella* spp. in Australia and provided further evidence of genomic evolution within the genus.

The *B. pertussis* genomes sequenced contained representatives of the predominating strain (*ptxP3*) in NSW with one being PRN-positive (CIDM-BP5) and one PRN-negative (CIDM-BP2), and a unique *ptxP1* PRN-positive isolate from 2015 (CIDM-BP3). *In silico* analysis confirmed that CIDM-BP3 and CIDM-BP5 each harboured an intact *prn* gene with no evidence of IS*481* insertion, supporting their phenotypic PRN production[21]. The *prn* gene in CIDM-BP2 contained an IS*481* insertion in position 1598, thus rendering the *prn* gene defective. These isolates were also classified with a previously described SNP clustering system, CIDM-BP2 and CIDM-BP5 are SNP Cluster I EL1 as SP12 and SP13 respectively, and CIDM-BP3 is a non-Cluster I SP18[21]. The isolates were also screened for known resistance mechanisms as macrolide resistant strains of *B. pertussis* have been observed in China. Macrolide resistance is the result of a 23S ribosomal RNA SNP mutation[25, 26] and none of the study isolates exhibited genotypes consistent with resistance in this gene.

Australian *B. pertussis* isolates harbour genomic rearrangements similar to those described previously in the northern hemisphere[14, 15]. Utilising the same clustering system, CIDM-BP3

and CIDM-BP5 were found to be singletons and CIDM-BP2 was part of Cluster-BP-12. For *B. parapertussis,* CIDM-BPP2 aligned closest with Cluster-PP-02 and singleton B160. However, with a deletion of a 1.6kb region in Cluster-PP-02, it supports the addition of CIDM-BPP2 to singleton B160, and the formation of a new cluster. CIDM-BH3 was more closely related to other closed *B. holmesii* isolates from the U.S.A and clustered with Cluster-BH-01. Interestingly, core genome phylogeny can predict structural clusters, however, full confirmation of genome rearrangements is only possible with closed genome assemblies. Phylogenetic analysis on its own goes some way to predict the structural cluster of draft genomes, however full confirmation requires long-read sequencing at a cost that is significantly higher than short-read high throughput sequencing.

The genomic inversions reported previously[14, 15] here, can be mediated by insertion sequences, primarily *IS481* in *B. pertussis* and *B. holmesii*, and *IS1001* in *B. parapertussis*. As in previous work, the study found that *IS481* and *IS1001* flanked all ends of all identified rearrangements. These IS are present in multiple copies in all *Bordetella* genomes and can number at over 200 copies in *B. pertussis*. The impact of these IS elements requires further investigation as previous studies have indicated that IS elements can upregulate upstream and downstream flanking gene expression in *B. pertussis*[27]. This lends to the hypothesis that specific structural clusters could be more virulent than others[27], as in some cases, *IS481* was introduced in front of transcriptional regulators post-inversion. The hypothesis can be addressed using RNA-Seq to resolve the transcriptome of strains from different structural clusters and comparing the impact of gene expression due to *IS481*.

A deeper core genome analysis into *B. holmesii* isolates in NSW showed that between 1999 and 2012, isolates were part of a specific cluster, differing by up to 32 SNPs only. More recent isolates (post-2012), however, differ from pre-2012 isolates by more than 72 SNPs indicating a potential change in circulating strains. Although, the small number of *B. holmesii* isolates in NSW makes a solid conclusion hard to draw. Given the more recent discovery of *B. holmesii* causing pertussis-like symptoms, a culture-independent approach to sequencing can be applied to *B. holmesii* positive respiratory samples to potentially close this gap in knowledge of the molecular epidemiology of *B. holmesii* in Australia.

Virulence marker investigations across the *Bordetella* spp. revealed that various virulence-associated genes of *B. pertusiss,* were not present on the genome of *B. parapertussis* or *B. holmesii*. In *B. parapertussis* CIDM-BPP2, all components of a T6SS were discovered, homologous to the T6SS found in *P. aeruginosa* and *V. cholerae*. This T6SS has also been discovered in *B. bronchiseptica* RB50, and whether the T6SS in *B. parapertussis* remains functional will have to be further investigated. *B. parapertussis* carries all the same virulence genes as *B. pertussis* with varying mutations including the *ptx-ptl* operon, however, as previously demonstrated, this operon in *B. parapertussis* is dysfunctional[28]. In comparison, *B. holmesii* lacks a large majority (87%) of the virulence markers found in *B. pertussis,* these include the vaccine antigens of *B. pertussis*.

In conclusion, the study applied structural clustering on closed genomes of the *Bordetella* spp. to determine where NSW isolates sit on a global scale. This study also presents the first five high-quality Southern hemisphere closed genomes of *Bordetella* spp. These reference genomes are crucial for future public health investigations of this significant group of

respiratory pathogens given the level of genomic rearrangement displayed by all species investigated in this study. While this study demonstrates further evidence of genomic evolution of the *Bordetella* species, dwindling isolate numbers due to diagnostic testing changes hamper the capability to continue monitoring *Bordetella* spp.

Methods

Strain selection and culture conditions

Chosen strains (Table 3) were stored at the Centre of Infectious Diseases and Microbiology Laboratory Services (CIDMLS) Identification Laboratory were cultured on Charcoal Blood Agar (*B. pertussis* and *B. parapertussis*) without Cephalexin (CBA) or Horse Blood Agar (HBA) (*B. holmesii*) for 7 days at 37°C.

Table 3: *Bordetella* spp. strains selected for both Nanopore and Illumina sequencing

Strains	Isolation Year	Isolation Site	Sequencing
Bordetella pertussis			
CIDM-BP2	2011	Respiratory	Nanopore
CIDM-BP3†	2015	Respiratory	Nanopore
CIDM-BP5	2008	Respiratory	Nanopore
Bordetella parapertussis			
CIDM-BPP2	1993	Respiratory	Nanopore
Bordetella holmesii			
CIDM-BH1	2000	Blood	Illumina
CIDM-BH2	2014	Blood	Illumina
CIDM-BH3	2016	Blood	Nanopore
CIDM-BH4	2000	Blood	Illumina
CIDM-BH6	1999	Blood	Illumina
CIDM-BH7	2007	Unknown	Illumina
CIDM-BH8	2012	Blood	Illumina

†This isolate has been published previously as L2263/BP7[21]

DNA Extraction and Quality Control Metrics

Two forms of DNA extraction were employed in this study, as not all strains were sequenced by Nanopore sequencing on the MinION platform (Table 3). DNA was extracted from isolates for long read sequencing (Nanopore) using the DNeasy® UltraClean® Microbial Kit (Qiagen, Germany) with modifications. Briefly, two loopfuls of culture were picked and inoculated into the Powerbead tube provided containing both Powerbead solution and solution SL. Cells were lysed mechanically for 2 minutes and subsequent steps were performed according to Manufacturer's instructions. DNA integrity was assessed by running the extracted DNA on a 0.6% (w/v) agarose gel for 3 hours at 75V in 1X TBE sized with a Quick-Load® 1 kb Extend DNA ladder (New England Biosciences, USA) to ensure that majority of the fragment sizes were > 30 kb.

Isolates for short read sequencing (Illumina) were extracted with the DNeasy Blood and Tissue kit (QIAGEN, Germany) with modifications. Two colonies were picked from the culture plates and inoculated directly into the buffer supplied in the kit and Proteinase K. The cells were lysed at 56°C on a heater-shaker for 3 hours, and subsequently extracted following the manufacturer's instructions.

All extracted DNA were quality assessed spectrophotometrically (AllSheng, China) and DNA quantity for long read sequencing was measured on a Qubit™ 2.0 Fluorimeter using the dsDNA BR Assay Kit (ThermoFisher Scientific, USA). DNA quantity for short read sequencing

was measured using Quant-iT™ PicoGreen™ dsDNA Assay Kit (ThermoFisher Scientific, USA).

DNA was stored in -20°C until library preparation.

Library Preparation and Sequencing

All extracted DNA was prepared with the Illumina NexteraXT DNA Library Preparation Kit v2.5 (Illumina, USA) and sequenced on the Illumina NextSeq 500.

Libraries for isolates that were sequenced on the Nanopore was prepared using the Rapid barcoding Kit SQK- RBK 004 (Oxford Nanopore, U.K) using > 1,000 ng of input DNA. Apart from the change in input amount, all subsequent manufacturer's instructions were followed. DNA was sequenced on a FLO-MIN106 flow cell for 24-48 hours without basecalling.

Bioinformatics

Base calling, assembly and annotation

Long read base calling in high accuracy mode and demultiplexing was performed on Guppy (v 2.4.5) on a GPU Amazon Web Service instance (c5.Metal). Demultiplexed reads were then _de novo_ assembled with Flye (v 2.7b)[29] with the "--asm-coverage" parameter set to 30 and an expected genome size of 4.0 Mb. Following long-read assembly, the sequence was corrected with Racon (v 1.3.1)[30] four times, and Medaka (v 0.11.5)[30] twice. The assembly was then polished with corresponding Illumina reads using Pilon (v 1.23)[31] and repeated until there were no more changes. The assembly was subsequently reordered so that the open reading frame of _dnaA_ was in the positive strand.

All short read sequences were assembled with SPAdes (v 3.12.0)[32], then annotated with Prokka (v 1.12)[33] and Barnapp (v 0.6) (https://github.com/tseemann/barrnap), then scanned for virulence factors (VFDB)[34] and resistance markers (CARD)[35] with ABRicate (v 0.9.8) (https://github.com/tseemann/abricate). The type VI secretions system analysis was performed by SecReT6[36].

Phylogenetic analysis

To determine where the Australian *Bordetella* spp. isolates sat globally, all closed RefSeq genomes were downloaded. Raw reads from SRA were also utilised and assembled with SPAdes (v 3.12.0)[32]. All assemblies and closed sequences were annotated with Prokka (v 1.12)[33] and Barnapp (v 0.6) (https://github.com/tseemann/barrnap).

For *B. pertussis,* a random selection of RefSeq sequences were extracted based on country and year of isolation (n=189) In addition, previously described closed genomes[14] were downloaded from RefSeq (n=472) and reanalysed with Australian *B. pertussis* isolates. SNP clusters were determined by BLASTn[37] with a custom database of SNPs positions from a previously published study[10, 11]. *B. holmesii* analysis was performed with all available RefSeq (n=84) and SRA sequences (n=10). SNP analysis of *B. holmesii* was performed by Snippy (v 4.1.0) and SNP-dist (v 0.6). For *B. parapertussis*, only closed sequences featured in Weigand et al were used for analysis. A list of all strains used in the study are in Supplementary Material 1.

For each *Bordetella* spp. generated GFF files from Prokka were used for phylogenetic analysis with Roary (v 3.1.2)[38], and the tree was drawn with IQ-TREE (v 1.6.7)[39] and

ModelFinder[40], with the best model being TVM+F+I. The core phylogenetic tree was

visualised with iTOL (v 5.7)[41]. Utilising the phylogenetic results, strains closest to the

Australian isolates were extracted for structural rearrangement analysis with

progressiveMauve (v 2.4.0)[22], the same cut-offs applied by Weigand et al[14] were applied in

the study.

Genome sequencing data has been published on NCBI under Bioproject: PRJNA695314

References

1.   Mattoo S, Cherry JD. Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to *Bordetella pertussis* and other *Bordetella* subspecies. Clin Microbiol Rev. 2005;18(2):326-82. doi:10.1128/CMR.18.2.326-382.2005

2.   Fong W, Timms V, Holmes N, Sintchenko V. Detection and incidence of *Bordetella holmesii* in respiratory specimens from patients with pertussis-like symptoms in New South Wales, Australia. Pathology. 2018;50(3):322-6. doi:10.1016/j.pathol.2017.10.014

3.   Pittet LF, Emonet S, Schrenzel J, Siegrist CA, Posfay-Barbe KM. *Bordetella holmesii*: an under-recognised *Bordetella* species. Lancet Infect Dis. 2014;14(6):510-9. doi:10.1016/S1473-3099(14)70021-0

4.   Lam C, Octavia S, Ricafort L, Sintchenko V, Gilbert GL, Wood N, et al. Rapid increase in pertactin-deficient *Bordetella pertussis* isolates, Australia. Emerg Infect Dis. 2014;20(4):626-33. doi:10.3201/eid2004.131478

5.   Safarchi A, Octavia S, Luu LD, Tay CY, Sintchenko V, Wood N, et al. Pertactin negative *Bordetella pertussis* demonstrates higher fitness under vaccine selection pressure in a mixed infection model. Vaccine. 2015;33(46):6277-81. doi:10.1016/j.vaccine.2015.09.064

6.   Weigand MR, Pawloski L, Peng Y, Ju H, Burroughs M, Cassiday PK, et al. Screening and genomic characterisation of filamentous haemagglutatin-deficient *Bordetella pertussis*. Infect Immun. 2018.

7.   Williams MM, Sen K, Weigand MR, Skoff TH, Cunningham VA, Halse TA, et al. *Bordetella pertussis* Strain Lacking Pertactin and Pertussis Toxin. Emerg Infect Dis. 2016;22(2):319-22. doi:10.3201/eid2202.151332

8.   Mooi FR, van Loo IHM, van Gent M, He Q, Bart MJ, Heuvelman K, et al. *Bordetella pertussis* Strains with Increased Toxin Production Associated with Pertussis Resurgence. Emerg Infect Dis. 2009;15(8):1206-13. doi:0.3201/eid1508.081511

9.   Bouchez V, Guglielmini J, Dazas M, Landier A, Toubiana J, Guillot S, et al. Genomic Sequencing of *Bordetella pertussis* for Epidemiology and Global Surveillance of Whooping Cough. Emerg Infect Dis. 2018;24(6):988-94. doi:10.3201/eid2406.171464

10.   Octavia S, Maharjan RP, Sintchenko V, Stevenson G, Reeves PR, Gilbert GL, et al. Insight into evolution of *Bordetella pertussis* from comparative genomic analysis: evidence of vaccine-driven selection. Mol Biol Evol. 2011;28(1):707-15. doi:10.1093/molbev/msq245

11.   Safarchi A, Octavia S, Wu SZ, Kaur S, Sintchenko V, Gilbert GL, et al. Genomic dissection of Australian *Bordetella pertussis* isolates from the 2008-2012 epidemic. J Infect. 2016;72(4):468-77. doi:10.1016/j.jinf.2016.01.005

12.	Alkan C, Sajjadian S, Eichler EE. Limitations of next-generation genome sequence assembly. Nat Methods. 2011;8(1):61-5. doi:10.1038/nmeth.1527

13.	Register KB, Sanden GN. Prevalence and sequence variants of *IS481* in *Bordetella bronchiseptica*: implications for *IS481*-based detection of *Bordetella pertussis*. J Clin Microbiol. 2006;44(12):4577-83. doi:10.1128/JCM.01295-06

14.	Weigand MR, Peng Y, Loparev V, Batra D, Bowden KE, Burroughs M, et al. The History of *Bordetella pertussis* Genome Evolution Includes Structural Rearrangement. J Bacteriol. 2017;199(8). doi:10.1128/JB.00806-16

15.	Weigand MR, Peng Y, Batra D, Burroughs M, Davis JK, Knipe K, et al. Conserved Patterns of Symmetric Inversion in the Genome Evolution of *Bordetella* Respiratory Pathogens. mSystems. 2019;4(6). doi:10.1128/mSystems.00702-19

16.	Olson ND, Lund SP, Colman RE, Foster JT, Sahl JW, Schupp JM, et al. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Front Genet. 2015;6:235. doi:10.3389/fgene.2015.00235

17.	Baines SL, da Silva AG, Carter GP, Jennison A, Rathnayake I, Graham RM, et al. Complete microbial genomes for public health in Australia and the Southwest Pacific. Microb Genom. 2020. doi:10.1099/mgen.0.000471

18.	Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015;12(8):733-5. doi:10.1038/nmeth.3444

19.	Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom. 2017;3(10):e000132. doi:10.1099/mgen.0.000132

20.	Ring N, Abrahams JS, Jain M, Olsen H, Preston A, Bagby S. Resolving the complex *Bordetella pertussis* genome using barcoded nanopore sequencing. Microb Genom. 2018;4(11). doi:10.1099/mgen.0.000234

21.	Xu Z, Octavia S, Luu LDW, Payne M, Timms V, Tay CY, et al. Pertactin-Negative and Filamentous Hemagglutinin-Negative *Bordetella pertussis*, Australia, 2013-2017. Emerg Infect Dis. 2019;25(6):1196-9. doi:10.3201/eid2506.180240

22.	Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010;5(6):e11147. doi:10.1371/journal.pone.0011147

23.	Allen AG, Thomas RM, Cadisch JT, Maskell DJ. Molecular and functional analysis of the lipopolysaccharide biosynthesis locus wlb from *Bordetella pertussis, Bordetella parapertussis* and *Bordetella bronchiseptica*. Mol Microbiol. 1998;29(1):27-38. doi:10.1046/j.1365-2958.1998.00878.x

24.	Bart MJ, van Gent M, van der Heide HG, Boekhorst J, Hermans P, Parkhill J, et al. Comparative genomics of prevaccination and modern *Bordetella pertussis* strains. BMC Genomics. 2010;11:627. doi:10.1186/1471-2164-11-627

25.   Wang Z, Cui Z, Li Y, Hou T, Liu X, Xi Y, et al. High prevalence of erythromycin-resistant *Bordetella pertussis* in Xi'an, China. Clin Microbiol Infect. 2014;20(11):O825-30. doi:10.1111/1469-0691.12671

26.   Xu Z, Wang Z, Luan Y, Li Y, Liu X, Peng X, et al. Genomic epidemiology of erythromycin-resistant *Bordetella pertussis* in China. Emerg Microbes Infect. 2019;8(1):461-70. doi:10.1080/22221751.2019.1587315

27.   Amman F, D'Halluin A, Antoine R, Huot L, Bibova I, Keidel K, et al. Primary transcriptome analysis reveals importance of IS elements for the shaping of the transcriptional landscape of *Bordetella pertussis*. RNA Biol. 2018;15(7):967-75. doi:10.1080/15476286.2018.1462655

28.   Hausman SZ, Cherry JD, Heininger U, Wirsing von Konig CH, Burns DL. Analysis of proteins encoded by the ptx and ptl genes of *Bordetella bronchiseptica* and *Bordetella parapertussis*. Infect Immun. 1996;64(10):4020-6. doi:10.1128/IAI.64.10.4020-4026.1996

29.   Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540-6. doi:10.1038/s41587-019-0072-8

30.   Vaser R, Sovic I, Nagarajan N, Sikic M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27(5):737-46. doi:10.1101/gr.214270.116

31.   Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9(11):e112963. doi:10.1371/journal.pone.0112963

32.   Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77. doi:10.1089/cmb.2012.0021

33.   Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9. doi:10.1093/bioinformatics/btu153

34.   Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 2005;33(Database issue):D325-8. doi:10.1093/nar/gki008

35.   McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother. 2013;57(7):3348-57. doi:10.1128/AAC.00419-13

36.   Li J, Yao Y, Xu HH, Hao L, Deng Z, Rajakumar K, et al. SecReT6: a web-based resource for type VI secretion systems found in bacteria. Environ Microbiol. 2015;17(7):2196-202. doi:10.1111/1462-2920.12794

37.   Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. doi:10.1186/1471-2105-10-421

38.   Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691-3. doi:10.1093/bioinformatics/btv421

39.   Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268-74. doi:10.1093/molbev/msu300

40.   Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-9. doi:10.1038/nmeth.4285

41.   Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007;23(1):127-8. doi:10.1093/bioinformatics/btl529