MICROBIAL GENOMICS

RESEARCH ARTICLE

Fong *et al.*, *Microbial Genomics* 2020;6

DOI 10.1099/mgen.0.000332

# Optimization of sample preparation for culture-independent sequencing of *Bordetella pertussis*

Winkie Fong[1,2,3,*], Rebecca Rockett[1,2], Verlaine Timms[1,2]† and Vitali Sintchenko[1,2,3]†

**Abstract**

*Bordetella pertussis*, the aetiological agent of whooping cough, is re-emerging globally despite widespread vaccination. *B. pertussis* is highly infectious and, prior to vaccination programmes, was the leading cause of infant mortality. The WHO estimated that over 600000 deaths are prevented annually by pertussis vaccination, but *B. pertussis* infection was still responsible for over 63000 deaths globally in 2013. The re-emergence of *B. pertussis* has been linked to strains with inactive or absent major virulence factors included in vaccines such as pertactin, pertussis toxin and filamentous haemagglutinin. Thus, the molecular surveillance of currently circulating strains is critical in understanding and controlling *B. pertussis*. Such information provides data on strains to inform control measures and the identification of future vaccine antigens. Current surveillance and typing methods for *B. pertussis* rely on the availability of clinical isolates. However, since the 1990s, the majority of pertussis cases have been diagnosed by PCR, where an isolate is not needed. The rapid decline in the availability of *B. pertussis* isolates impacts our ability to monitor this infection. The growing uptake of next-generation sequencing (NGS) has offered the opportunity for culture-independent genome sequencing and typing of this fastidious pathogen. Therefore, the objective of the study was to optimize respiratory sample preparation, independent of culture, in order to type *B. pertussis* using NGS. The study compared commercial depletion kits and specimen-processing methods using selective lysis detergents. The goal was to deplete human DNA, the major obstacle for sequencing a pathogen directly from a clinical sample. Samples spiked with a clinically relevant amount of *B. pertussis* were used to provide comparison between the different methods. Commercial depletion kits including the MolYsis, Qiagen Microbiome and NEBNext Kits were tested. Previously published methods, for Saponin and TritonX-100, were also trialled as a depletion. The ratio of *B. pertussis* to human DNA was determined by real-time PCR for ERV3 and *IS481* (as markers of human and *B. pertussis* DNA, respectively), then samples were sequenced using the Illumina NextSeq 500 platform. The number of human and *B. pertussis* sequenced reads were then compared between treatments. The results showed that commercial kits reduced the human DNA present, but also reduced the concentration of target *B. pertussis*. However, selective lysis with Saponin treatment resulted in almost undetectable levels of human DNA, with minimal loss of target *B. pertussis* DNA. Sequencing read depth improved five-fold in reads to *B. pertussis*. Our investigation delivered a potential protocol that will enable the public health laboratory surveillance of *B. pertussis* in the era of culture-independent testing.

## DATA SUMMARY

The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files (available in the online version of this article). Reads from mapping to NC_002929.2 *B. pertussis* Tohama I have been uploaded to SRA under BioProject: PRJNA595393.

## INTRODUCTION

Pertussis, or whooping cough, is an acute and highly transmissible respiratory infection, associated with prolonged coughing episodes and mortality [1]. While symptoms can be mild in adults, the infection is life-threatening in infants and the elderly [1]. *Bordetella pertussis*, the main aetiological

Received 21 August 2019; Accepted 09 January 2020; Published 28 February 2020

**Author affiliations:** [1]Centre for Infectious Diseases and Microbiology – Public Health, Westmead Hospital, Westmead, NSW, Australia; [2]Westmead Clinical School, The University of Sydney, Westmead, NSW, Australia; [3]Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney, Camperdown, NSW, Australia.

***Correspondence:** Winkie Fong, wfon4473@uni.sydney.edu.au

**Keywords:** *Bordetella pertussis*; laboratory diagnosis; whole genome sequencing; microbial genomics.

**Abbreviations:** CIDMLS, Centre of Infectious Diseases and Microbiology Laboratory Services; FHA, filamentous haemagglutinin; HPC, high performance computing cluster; MLST, multi-locus sequence typing; MLVA, multiple locus variable-number tandem repeat analysis; NGS, next generation sequencing; NPA, nasopharyngeal aspirate; PRN, pertactin; PTX, pertussis toxin; rtPCR, real-time PCR.

†These authors share senior authorship.

**Data statement:** All supporting data, code and protocols have been provided within the article or through supplementary data files. Supplementary material is available with the online version of this article.

agent of whooping cough, is highly infectious and, despite widespread vaccination, is re-emerging as a significant pathogen globally [2]. The disease is primarily controlled by immunization, as antibiotics alone do not protect vulnerable populations such as newborns and the elderly [3]. In the years following the introduction of pertussis vaccination in Australia in the 1960s, there were on average 5000 notifications per year, whereas in 2011, there were over 38000 cases. The reasons for this increased incidence are unconfirmed, but may include increased testing, waning immunity or 'vaccine escape' of *B. pertussis* strains. Vaccine escape occurs when the pathogen mutates to avoid the vaccine-associated immune response. Evidence of vaccine pressure is demonstrated by genetic drift and variation of the vaccine antigens, such as the emergence of variants of pertussis toxin (PTX) promoters, as well as pertactin (PRN)-deficient [4, 5], filamentous haemagglutinin (FHA)-deficient [6] or PTX-deficient [7] strains of *B. pertussis*. The deficiency of these vaccine antigens allows evasion of the immune response provided by the vaccine, therefore rendering the world-wide vaccines less effective. However, vaccination remains the best method of pertussis control, given that antibiotic therapy is only effective during the early stage of disease [8]. Thus, surveillance of currently circulating strains becomes paramount to the control of *B. pertussis* in the community, particularly for the early detection of strains not covered by the vaccine.

However, molecular surveillance using traditional methods such as multiple locus variable-number tandem repeat analysis (MLVA) and more recently next generation sequencing (NGS) relies on a laboratory isolate, grown from a clinical sample from an infected patient. Culture isolates can be used in a range of laboratory tests for typing and antibiotic susceptibility assessment. The culture isolates can also be stored and resuscitated at a later date to examine past and current whooping cough epidemics. In the past, the culturing of an isolate was the primary laboratory method used to identify an infection and therefore provided ample material for public health surveillance. However, recent developments in laboratory diagnostics mean that a live isolate is no longer required and, today, *B. pertussis* is primarily detected in clinical samples by real-time PCR (rtPCR)-based assays targeting *IS481*. Other information such as strain type or antibiotic resistance cannot be readily determined from these molecular tests or the samples sent to the laboratory. For a re-emerging pathogen, such as *B. pertussis*, live cultures are needed to monitor changes contributing to its re-emergence. Fewer isolates are available for *B. pertussis* surveillance, and how the pathogen is evading vaccine-derived immunity remains therefore unclear. To obtain data previously derived from culture isolates, current NGS methods need modifications to allow genotyping of PCR-positive clinical samples that is relevant to *B. pertussis* surveillance.

In addition, the current PCR-based diagnostic assays targeting only *IS481* are unable to distinguish between *B. pertussis* and other clinically relevant *Bordetella* species which can co-circulate and be misidentified in pertussis epidemics. One such species, *B. holmesii*, is an emerging infection causing pertussis-like symptoms, mainly in the 11–18 year age group [9, 10], which also carries copies of the *IS481* element [11]. This affects the quality of public health laboratory surveillance data and also the assessment of vaccine efficacy for whooping cough.

**Impact Statement**

The rising incidence of *Bordetella pertussis* over the past decade highlights the critical need to understand how the pathogen evades vaccine-generated immunity. However, to fully assess the resurgence, high-resolution typing of *B. pertussis* is vital. Next generation sequencing (NGS) provides the ultimate resolution in bacterial typing, but it is currently reliant on bacterial isolation prior to sequencing. The diagnosis of *B. pertussis* is almost exclusively performed using molecular-based diagnostics, circumventing culture of this fastidious pathogen. As a result, there is a desperate need for culture-free typing methods. Utilizing NGS on clinical specimens has the potential to generate high-resolution typing information. The main disadvantage of NGS directly applied to clinical samples is the presence of human cells. These human cells contain DNA that overwhelms the sequencing process and dominates the output. Hence, depletion of human DNA prior to sequencing can potentially reduce its proportion in the final sequence data. We compared commercial and in-house depletion methods with the aim to reduce human DNA before sequencing. Our results show that Saponin-treated samples vastly increased the sequencing reads directed to *B. pertussis* and can be used in conjunction with deep sequencing to provide targeted culture-free high-resolution typing.

Culture-independent sequencing utilizes the high-resolution capability of NGS and applies it to clinical samples. Even though applying NGS directly to clinical samples has been shown to be technically challenging [12–14], it still offers the greatest potential for high-resolution data. The presence of human cells poses the greatest challenge for culture-independent sequencing of clinical samples. The human genome is 1000 times larger than a bacterial genome, and combined with low bacterial numbers, extracted human DNA can vastly exceed the target bacterial DNA [14]. Manual depletion of human DNA prior to sequencing experiments increases the chance of obtaining sequences from low-abundance pathogens. Several commercial kits have been designed to deplete human DNA from clinical samples for this purpose. The NEBNext Microbiome Kit targets the CpG-methylated nature of the human DNA, thereby selectively binding and removing it from the sample [15]. The MolYsis Basic Kit and QiaAMP Microbiome Kit selectively lyse human cells, leaving only bacterial cells intact for future lysis [16, 17]. Alternative selective lysis detergents have also been used such as TritonX-100 and Saponin.

This study describes and contrasts methods that optimize respiratory sample processing to enable NGS of bacterial

pathogens without the need for a bacterial culture. We focus on sample processing as previous reports have shown that culture-independent sequencing of the DNA extracted from clinical samples results in mainly human DNA [12]. The results presented in this study can form the basis for future sequencing and typing of low-load pathogens from respiratory tract samples.

## METHODS

### Specimens

Nasopharyngeal aspirates (NPAs) were obtained from the Centre of Infectious Diseases and Microbiology Laboratory Services (CIDMLS), NSW Health Pathology. CIDMLS provided specimens from 2017, which would otherwise have been discarded. These NPA specimens were *B. pertussis IS481*-negative by PCR. To reduce biological variation for this comparison, the specimens were then pooled into a tube and stored at −20 °C until further use. NPAs were spiked with 13600 c.f.u. *B. pertussis* ATCC9797 to mimic the levels of *B. pertussis* present in specimens obtained from patients with pertussis. This bacterial density estimate was based on average cycle threshold (Ct) values obtained from clinical *B. pertus*sis-positive samples tested in our previous study [10]. Spiked NPA was used instead of clinical NPA to enable comparison across different depletion protocols. By using the same pooled NPA, variation between samples was minimized. Clinical samples positive for *B. pertussis* may introduce excessive variation in *B. pertussis* numbers for accurate comparison of sample preparation methods.

### Human DNA depletion and target DNA extraction protocols

The spiked samples were then used to determine the best protocol for target microbial DNA extraction. Samples were processed in duplicate with a pair left untreated (without depletion) as a control. A total of three commercially available kits were tested: New England Biolabs' NEBNext Microbiome DNA Enrichment kit, Molzym's MolYsis Basic kit and Qiagen's QIAmp Microbiome DNA Kit. Two selective-lysis detergents, Saponin and TritonX-100, were also tested and compared to commercial kits.

#### NEBNext Microbiome Kit

The first protocol performed was with the NebNext Microbiome Kit, following the manufacturer's protocol.

#### MolYsis Basic Kit

The MolYsis Basic Kit was used according to the manufacturer's instructions without modifications. MolYsis treatments required a separate extraction protocol, for which we used extraction by the Qiagen DNeasy Blood and Tissue kit or the Promega Wizard Genomic Purification Kit.

#### Qiagen QIAamp DNA Microbiome Kit

The third protocol utilized the Qiagen QIAamp DNA Microbiome Kit, in accordance with the manufacturer's protocol.

#### Saponin and TritonX-100

The Saponin and TritonX-100 protocol was as follows: spiked NPA specimens were divided into 200 µl aliquots and mixed with 1% Saponin (Sigma Aldrich) or 1% TritonX-100 (Sigma Aldrich) to a final concentration of 0.025 or 0.012%. Samples were then vortexed for 10 s and incubated at room temperature for 5 min. This is followed by 10× Turbo DNase Buffer (ThermoFisher Scientific) addition to a final concentration of 1× and 4 U of Turbo DNase (ThermoFisher Scientific). The sample was then extracted with the Qiagen DNeasy Blood and Tissue kit, following the manufacturer's protocol.

### Measurement of genomic DNA concentration and quality

Genomic DNA concentration was measured via a Qubit 2.0 fluorometer (ThermoFisher Scientific) utilizing the Qubit dsDNA HS assay kit (ThermoFisher Scientific). An Allsheng Nano-300 Spectrophotometer (Allsheng) was used to measure the quality and purity of DNA for $A_{260}$:$A_{280}$ and $A_{260}$:$A_{230}$ ratios.

Samples were then assessed for their depletion trials by rtPCR. The rtPCR targets the *ERV3* gene, a retrovirus DNA inserted in the human genome passed down from human ancestors [18]. In addition, a second separate rtPCR targeted the *IS481* insertion sequence of *B. pertussis* to assess the depletion methods. Protocols were judged as successful when a low *IS481* Ct value compared to the ERV3 Ct was obtained. This indicated low human DNA, and high *B. pertussis*. Analysis was made in comparison to an undepleted sample from the same NPA pool.

The *ERV3* PCR assay included the following; 250 nM for *ERV3*19 reverse primers each (Sigma Aldrich), 100 nM TaqMan probe (Sigma Aldrich), 1× Roche LightCycler ProbeMaster (Roche Diagnostics), and molecular-grade PCR water to make a volume of 18 µl mastermix. Then, 2 µl of DNA extract was added, resulting in a total reaction volume of 20 µl.

The *IS481* PCR mixture (20 µl) contained QIAGEN HotStarTaq MasterMix (Qiagen), made up to final concentrations of 1.5 mM $MgCl_2$, 200 µM dNTPs and 1.23 U HotStarTaq DNA polymerase (Qiangen), 300 nM of *IS481* forward and *IS481* reverse primers (Sigma Aldrich), 120 nM TaqMan FAM probe (Sigma Aldrich) and 5 µl DNA extract.

The thermocycling profile for the amplification and detection of both the *IS481* and *ERV3* assays consisted of a denaturation step of 95 °C for 10 min, followed by 45 cycles of amplification at 95 °C for 10 s, 55 °C for 30 s and 72 °C for 15 s, at a ramp rate of 4.4 °C $s^{-1}$, and the reaction was allowed to cool at 40 °C for 10 min.

Samples were forwarded for sequencing if they surpassed an rtPCR *IS481* Ct cut-off of 37 cycles. Furthermore, only an *IS481* rtPCR was used to gauge the presence of *B. pertussis,* as NPA was spiked with a known *B. pertussis* isolate.

### Library preparation and sequencing

The sequencing library was prepared by using the NexteraXT DNA Sample preparation kit (Illumina) following

the manufacturer's protocol, and was performed by the Pathogen Genomics Team at CIDM-PH. The samples were then sequenced on the Illumina NextSeq 500 platform on Mid-Throughput flow cells to achieve a coverage of 67–70× depth across all samples.

### Analysis of sequences

Following sequencing of all spiked and clinical NPAs, FASTQ files were downloaded from Illumina Basespace (Illumina) and transferred to the University of Sydney High Performance Computing Cluster (HPC). Analysis began by trimming the ends of poor-quality reads with a phred quality score of <15 with Trimmomatic. These trimmed reads were then mapped to the human genome GRCh38.p12 (GCA_000001405.27), and *B. pertussis* Tohama I reference genome (NC_002929.2) by Burrows-Wheeler Aligner (BWA) (version 0.7.17). The number of reads mapped to the respective genomes were then used to determine the percentage of representative organisms by dividing the result of each of the mappings by the number of reads of the total sample. SAM and BAM files produced by BWA were then imported into the Qiagen CLC Genomics Workbench, and observations were made at regions of interests such as the 16S rRNA gene, multilocus sequence type (MLST) genes, and *ptx*, *prn* and *fhaB* encoding regions to determine coverage of the area. In addition, a detailed mapping report was generated to determine the number of reads mapped to the *B. pertussis* and human genomes.

## RESULTS

### Comparison of host DNA depletion methods

Each sample was homogenized and divided into volumes required by the commercial kit and processed in duplicate and in parallel. The first kit trialled was the NEBNext Microbiome enrichment kit, which reduced human DNA from a Ct value of 26.50 to 36.81, which corresponded to a 3.3 log loss of human DNA copies. The method did not influence the concentration of *B. pertussis*, which remained relatively constant at Ct values of 31.33–32.52. Upon sequencing, post-unspiked samples did not contain a single *B. pertussis* sequencing read as expected, while post-depletion samples resulted in 0.073% of reads being mapped to *B. pertussis*. These reads mapped to *B. pertussis* covered 2.85% of the genome.

The second kit assessed was the MolYsis Basic Kit, and the depleted samples were extracted by either Qiagen or Promega kits, then assessed by rtPCR targeting both *IS481* and *ERV3*. As shown in Table 1, the Qiagen extraction, when compared to an undepleted control, resulted in a reduction of human DNA from a Ct value of 22.88 to 35.63, which is a 4.0 log loss of human DNA copies. The MolYsis depletion, however, also reduced the concentration of *B. pertussis* during the treatment process, reducing the Ct value from 27.27 to 31.93, a 1.3 log loss of *B. pertussis* DNA. As *B. pertussis* was still detectable by PCR, these samples were then sequenced. The average percentage of human reads in the undepleted samples (92.5±0.8%) was reduced by almost half (58.4±13.1%). For *B. pertussis,* depletion was able to improve the percentage of reads 4-fold (0.61±0.05 to 2.12±0.40%) (Figs 1 and 2) and covered 5.55±1.29% of the genome.

With the aim to improve DNA yield, the Promega Wizard Genomic Purification Kit was trialled as the extraction method post-MolYsis treatment. PCR showed that compared to the Qiagen method, the Promega Wizard kit resulted in a reduced DNA concentration. Regardless, this method still demonstrated a reduction in human DNA post-MolYsis, shifting from a Ct value of 34.56 to being undetected by PCR. However, *B. pertussis* DNA was also reduced from 29.41 to 35.01, a 2 log loss. This was reflected in the overall reads recovered from sequencing (971478±69600 reads), almost half the number compared to the protocol with a Qiagen extraction (1855289±1021291 reads). However, the average percentage of human reads in the raw sample (69.3±1.7%) was significantly decreased (23.7±0.2%) overall. The number of reads mapped to the *B. pertussis* genome was similar to the Qiagen-extracted method (3.5±0.2%) and covered 5.7±0.1% of the genome.

The third commercial kit examined was the QIAamp DNA Microbiome Kit. PCR results indicated a reduced DNA concentration compared to the MolYsis Basic Kit. Human DNA had a Ct value of 35.83±0.23, while *B. pertussis* Ct values were 37.93±2.07. Hence due to the low bacterial concentration by rtPCR, these samples did not proceed to sequencing.

The depletion method, 0.025% Saponin, showed that compared to the undepleted samples, there was a reduction in human DNA concentration, shifting from a Ct value of 22.88±0.15 to undetectable by PCR, and only reducing *B. pertussis* DNA from 27.27±0.09 to 30.51±0.43. Thus, the method demonstrated limited impact of the detergent on the bacteria of interest while reducing the amount of human DNA. Sequencing statistics showed that 3.8±0.1% of the 2.7 million reads were to *B. pertussis*, and 12.9±0.5% were to the human genome. The *B. pertussis* reads covered 9.3±0.1% of the genome.

The application of TritonX-100 at a concentration of 0.025% significantly reduced the load of human DNA, although it came at the consequence of losing almost 3.3 log *B. pertussis* DNA (36.70±0.59). Sequencing yielded an average of 1157281±45617 reads for samples treated with TritonX-100. Despite the loss of overall reads, 2.24±0.06% of the reads were *B. pertussis* and covered 3.62±0.05% of the *B. pertussis* genome.

### Optimization of protocols and target DNA quality metrics

Observations with both Saponin and TritonX-100 protocols led to trialling both Saponin and TritonX-100 together to determine if lower concentrations of TritonX-100 could further reduce human DNA in comparison with Saponin alone. Four different ratios of Saponin and TritonX-100 were tested, 0.006 or 0.012% respectively, as outlined in Table 1. For the 0.012% Saponin and 0.012% TritonX-100 combination, PCR revealed human DNA was reduced significantly (38.12 and undetectable), but to similar levels

**Table 1.** Summary of depletion methods results

| Depletion method | PCR | | Sequencing | | | |
|---|---|---|---|---|---|---|
| | *IS481* | *ERV3* | % of *B. pertussis* reads | % of Human reads | Average coverage | Coverage of the *B. pertussis* genome (%) |
| Undepleted | | | | | | |
| Qiagen | 27.27 | 22.88 | 0.608 | 92.5 | 0.18 | 0.11 |
| Promega | 29.41 | 34.26 | 1.335 | 69.3 | 0.12 | 0.12 |
| Pre-NEBNext | 31.33 | 26.50 | 0.000 | 98.7 | 0 | 0.13 |
| Depleted | | | | | | |
| MolYsis/Qiagen | 31.93 | 35.63 | 2.115 | 58.4 | 0.28 | 5.55 |
| MolYsis/Promega | 35.01 | – | 3.449 | 23.7 | 0.27 | 5.72 |
| NEBNext | 32.52 | 36.81 | 0.073 | 98.3 | 0.06 | 2.85 |
| QIAamp | 37.93 | 35.83 | Not sequenced* | | | |
| 0.025% Saponin | 30.51 | – | 3.798 | 12.9 | 0.9 | 9.31 |
| 0.012% Saponin | 31.86 | 40.00 | 3.855 | 13.3 | 0.77 | 8.21 |
| 0.025% TritonX-100 | 36.70 | 38.70 | 2.237 | 39.2 | 0.22 | 3.62 |
| 0.012% TritonX-100 | 33.53 | 38.72 | 2.027 | 41.4 | 0.22 | 4.11 |
| 0.012% Saponin and 0.012% TritonX-100 | 36.70 | 38.12 | Not sequenced† | | | |
| 0.006% Saponin and 0.006% TritonX-100 | 34.04 | 38.01 | Not sequenced† | | | |
| 0.012% Saponin and 0.006% TritonX-100 | 34.55 | – | Not sequenced† | | | |
| 0.006% Saponin and 0.012% TritonX-100 | 35.96 | – | Not sequenced† | | | |

*Not sequenced due to low target recovery by rtPCR.
†Saponin and TritonX-100 combinations were not sequenced as they performed worse than Saponin-treated samples.

achieved by the TritonX-100-only depletion; *B. pertussis* DNA (36.70±0.45) was 1.6 log lower than with the Saponin-only-treated samples. The lower concentrations of Saponin and TritonX-100 (0.006% each) resulted in similar human DNA reductions (38.01±0.45) as with 0.012% Saponin and TritonX-100 combinations, with slightly higher concentrations of *B. pertussis* DNA (36.70±0.45). However, Ct cycles of these samples were still 1 log lower than in Saponin-only-treated samples. Sequencing was not performed on all trials, as based on PCR results, they performed significantly worse than Saponin-only-treated samples.

## DISCUSSION

As a result of transitions from culture-based to molecular-based diagnostics, sequencing of clinical samples without the need for culture can supply critical surveillance information for public health. Given that the human genome is significantly (1000×) larger than a bacterial genome, the concentration of human DNA vastly overwhelms the desired bacterial genome, with often >99% of sequencing reads from clinical samples being human DNA [20]. Methods over the last few years have attempted to reduce the human DNA prior to sequencing, allowing more bacterial DNA to be successfully sequenced.

In this study, we examined and contrasted several genomic DNA extraction and depletion methods in order to optimize sample preparation for direct sequencing of *B. pertussis*. Human DNA depletion was first attempted with the MolYsis Basic kit to process *B. pertussis*-spiked samples. Following depletion and extraction, the DNA concentration was assessed by rtPCR targeting the *ERV3* and *IS481* sequences. MolYsis treatment with extraction by the Qiagen Kit reduced human DNA to almost undetectable levels, but it also reduced the concentration of *B. pertussis* by nearly 2.6 log compared to the undepleted controls. Sequencing resulted in variable human to pertussis read percentages (human reads made up 40–70% of all reads); the cause of the variation could be due to GC bias of the sequencing method [21], which prioritizes

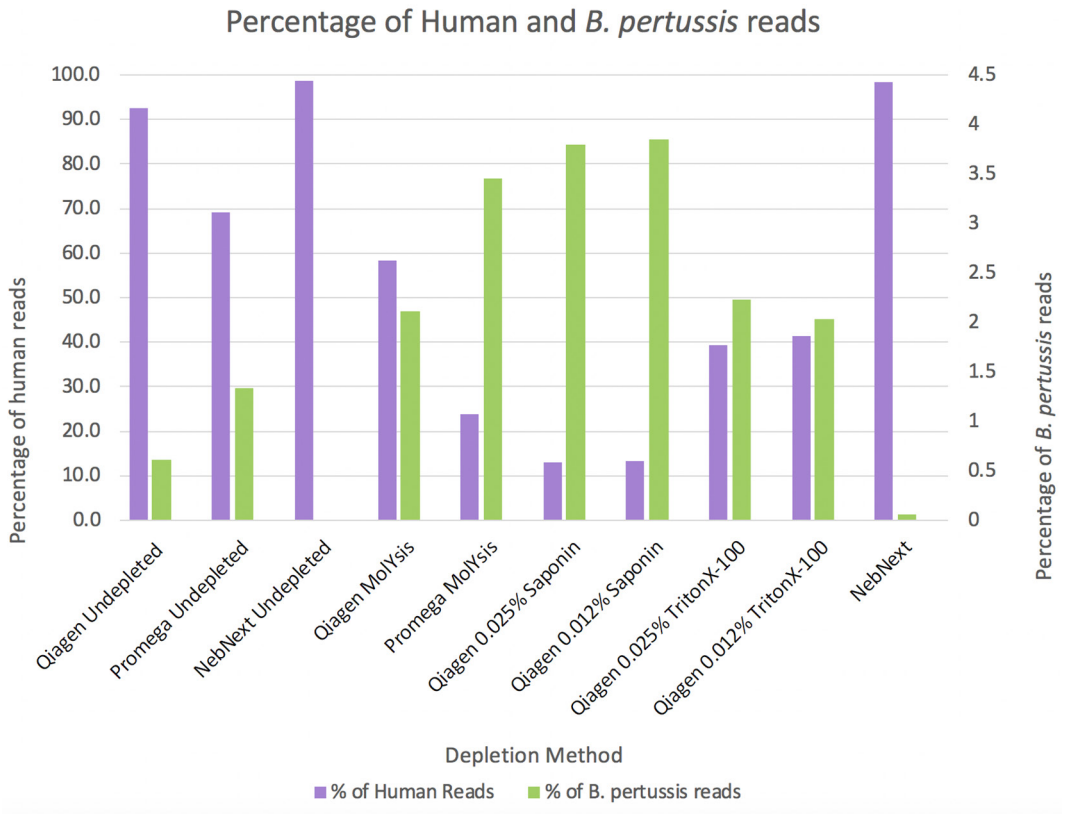


**Fig. 1.** Percentage of human (purple) and *B. pertussis* reads (green) compared to the total number of reads per sample. The first three bars represent the untreated samples, and each depletion extraction method was compared to its corresponding undepleted method (Note: Qiagen MolYsis was compared to Qiagen Undepleted).

GC% equal DNA, and therefore *B. pertussis* DNA with its GC content of 67% is at a disadvantage.

MolYsis-treated NPA was also extracted by the Promega kit to compare different DNA extraction methods, but this extraction procedure resulted in very low DNA concentrations for both human and *B. pertussis* DNA. MolYsis depletion extracted by the Promega kit lost a large proportion of DNA, potentially due to the nature of the extraction method, which had a step that required removal of the supernatant. The removal from an already low concentration of cells may have affected the visibility of the pellet, thus inadvertently removing DNA from the sample. Despite the low concentration of DNA, sequencing still progressed to determine baseline information regarding the sequencing limits of the technology. Regardless of the DNA concentration, MolYsis-depleted samples extracted by the Promega kit sequenced better than predicted, yielding about one million reads overall, although these still contained 23% human reads. A potential reason for this better than expected sequencing yield could be the low human DNA levels post-extraction, enabling increased sequencing efficiency of *B. pertussis* DNA.

An alternative commercial depletion kit, the Qiagen QIAamp DNA Microbiome, was again effective at removing human DNA, but rtPCR revealed a 3.3 log loss of *B. pertussis* within the sample. Because results were poorer than with the MolYsis-depleted Promega-extracted samples, these samples were not sequenced. Alternative methods were published by Hasan *et al.* utilizing Saponin and TritonX-100 as effective non-ionic detergents able to lyse human cells without interrupting bacterial cell walls [15]. This study was able to replicate the results from their study. For TritonX-100-treated samples, rtPCR demonstrated a significant loss in human DNA, although this protocol also significantly reduced *B. pertussis* DNA more than 3 log. Fortunately, selective lysis with Saponin resulted in an almost undetectable level of human DNA by rtPCR, with only a 2.6 log loss in *B. pertussis* DNA. When sequenced, Saponin-treated samples reduced the human DNA reads from 91% in the undepleted sample to 13%, a 10-fold reduction. This is in comparison to TritonX-100, which only reduced human reads to 50%, all while not improving *B. pertussis* reads. Hence in comparison to Saponin, TritonX-100 treatment did not perform well.

Sequencing of the DNA extracted with the above methods, showed low recovery of *B. pertussis* reads. However, the best pre-treatment method, 0.025% Saponin, demonstrated a 5-fold increase in the number of reads. Therefore, 0.025% Saponin treatment followed by Qiagen extraction was the best sample preparation method to recover *B. pertussis* DNA from NPA that is then suitable for sequencing.

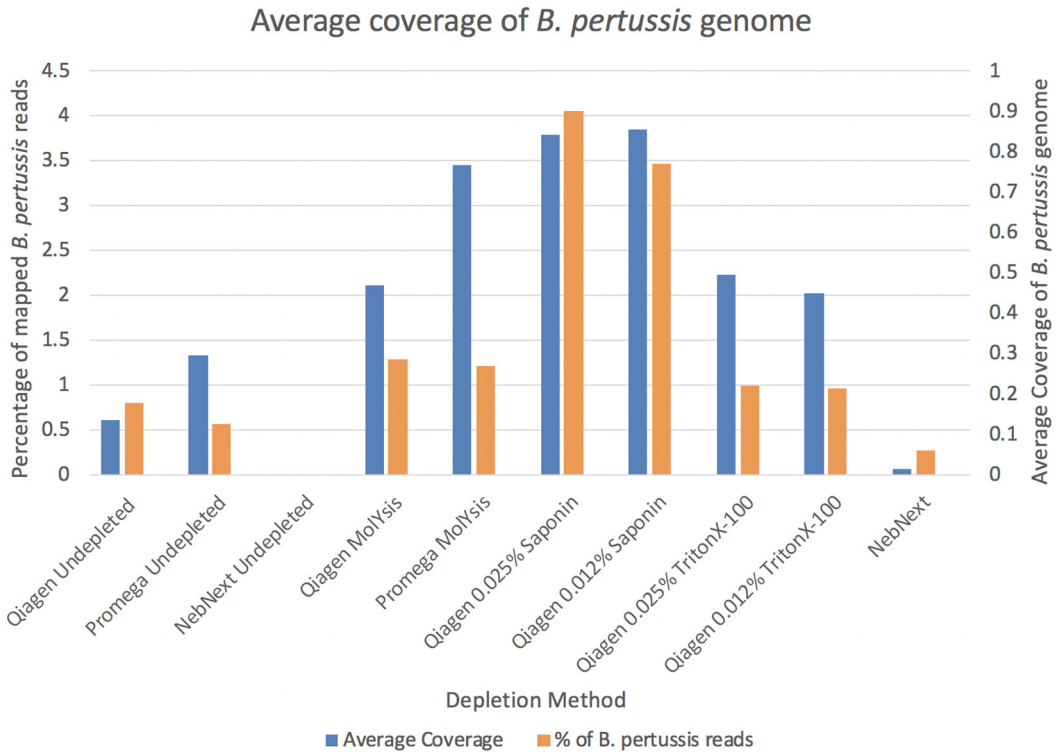


**Fig. 2.** Average coverage of the *B. pertussis* genome (blue) and percentage of mapped reads (orange) were plotted to show the efficiency of each depletion method. Saponin-treated methods improved data yield from sequencing methods.

Despite these promising results, we were unable to obtain information that distinguished allele types for MLST, nor determine gene alleles such as *ptxP*. In addition, the coverage of these genomes was insufficient to obtain a full genome, and complete gene sequences could not be recovered. Species differentiation between *B. pertussis* and *B. holmesii* was also difficult, as the 16S rRNA genes are highly conserved between the two species. However, the technology for direct sequencing of *B. pertussis* can be improved to increase depth and confidence of the data by gene-targeted enrichment [22, 23], or even deep sequencing [24].

In conclusion, we have demonstrated that commercial depletion kits and specimen-processing methods using selective lysis detergents can successfully deplete human DNA, the major obstacle for sequencing a pathogen directly from a clinical sample. Selective lysis with Saponin treatment can achieve almost undetectable levels of human DNA, with minimal loss of target *B. pertussis* DNA. Our findings provide an important first step in optimized sample preparation for future applications on clinical samples in place of isolates.

Funding information
This study was supported financially by the Centre for Infectious Diseases and Microbiology – Public Health Post-graduate scholarship

Acknowledgements
This work was supported by the Prevention Research Support Program, funded by the New South Wales Ministry of Health. The authors are grateful to the staff of the Centre for Infectious Diseases and Microbiology Laboratory Services, NSW Health Pathology, for their technical assistance and expertise. Computational analysis was performed on the University of Sydney HPC, with the assistance of the Sydney Informatics Hub. The authors would like to acknowledge Alicia Arnott, Nathan Bachmann, Chayanika Biswas, Rajat Dhakal, Elena Martinez, Ranjeeta Menon, Rosemarie Sadsad and Qinning Wang at the Pathogen Genomics Unit, Centre for Infectious Disease and Microbiology –Public Health, Westmead Hospital, for their assistance with genome sequencing and bioinformatics analysis.

Author contributions
The study was conceptualized by W.F., V.T. and V.S. Laboratory work and genome analysis was performed by W.F. The manuscript was written by W.F. and reviewed and edited by V.T., R.R. and V.S.

Conflicts of interest
The authors declare that there are no conflicts of interest

Ethical statement
NPAs were collected by the Centre for Infectious Diseases and Microbiology Laboratory services under the Western Sydney Local Health District Research Ethics and Governance committee. Project identifier: 2019/PID02294. Consent was not obtained from patients, as these NPAs were left over from previous diagnostic testing, and otherwise discarded. These NPA samples were pooled to prevent any identification during sequencing, and no record of identifiable data was collected.

Data bibliography
Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. doi:10.1093/bioinformatics/btu170.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. doi:10.1093/bioinformatics/btp324.

CLC Genomics Workbench 10.0 (https://www.qiagenbioinformatics.com/)

Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, *et al*. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat Genet. 2003;35(1):32-40. doi:10.1038/ng1227. GenBank BioProject PRJNA57617. Accession NC_002929.2 (2010).

Genome Reference Consortium Human Build 38 patch release 12 (GRCh38.p12). GenBank BioProject PRJNA31257. Accession GCA_000001405.27 (2017).

References

1. Stojanov S, Liese J, Belohradsky BH. Hospitalization and complications in children under 2 years of age with Bordetella pertussis infection. *Infection* 2000;28:106–110.
2. Crowcroft NS, Stein C, Duclos P, Birmingham M. How best to estimate the global burden of pertussis? *Lancet Infect Dis* 2003;3:413–418.
3. Cherry JD. Historical review of pertussis and the classical vaccine. *J Infect Dis* 1996;174:S259–S263.
4. Lam C, Octavia S, Ricafort L, Sintchenko V, Gilbert GL *et al*. Rapid increase in pertactin-deficient *Bordetella pertussis* isolates, Australia. *Emerg Infect Dis* 2014;20:626–633.
5. Safarchi A, Octavia S, Luu LDW, Tay CY, Sintchenko V *et al*. Pertactin negative Bordetella pertussis demonstrates higher fitness under vaccine selection pressure in a mixed infection model. *Vaccine* 2015;33:6277–6281.
6. Weigand MR, Pawloski LC, Peng Y, Ju H, Burroughs M *et al*. Screening and genomic characterization of filamentous hemagglutinin-deficient *Bordetella pertussis*. *Infect Immun* 2018;86.
7. Williams MM, Sen K, Weigand MR, Skoff TH, Cunningham VA *et al*. *Bordetella pertussis* Strain Lacking Pertactin and Pertussis Toxin. *Emerg Infect Dis* 2016;22:319–322.
8. Bass JW. Erythromycin for treatment and prevention of pertussis. *Pediatr Infect Dis J* 1986;5:154–157.
9. Rodgers L, Martin SW, Cohn A, Budd J, Marcon M *et al*. Epidemiologic and laboratory features of a large outbreak of pertussis-like illnesses associated with cocirculating Bordetella holmesii and Bordetella pertussis--Ohio, 2010-2011. *Clin Infect Dis* 2013;56:322–331.
10. 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;50:322–326.
11. Loeffelholz M, Thompson CJ, Long KS, Gilchrist MJR. Detection of *Bordetella holmesii* using *Bordetella pertussis* IS*481* PCR Assay. *J Clin Micro* 2000;38:467.
12. Yang J, Yang F, Ren L, Xiong Z, Wu Z *et al*. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. *J Clin Microbiol* 2011;49:3463–3469.
13. Gu W, Crawford ED, O'Donovan BD, Wilson MR, Chow ED *et al*. Depletion of abundant sequences by hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol* 2016;17:41.
14. Bachmann NL, Rockett RJ, Timms VJ, Sintchenko V. Advances in clinical sample preparation for identification and characterization of bacterial pathogens using Metagenomics. *Front Public Health* 2018;6:363.
15. Graham RMA, Doyle CJ, Jennison AV. Epidemiological typing of *Neisseria gonorrhoeae* and detection of markers associated with antimicrobial resistance directly from urine samples using next generation sequencing. *Sex Transm Infect* 2017;93:65–67.
16. Votintseva AA, Bradley P, Pankhurst L, Del Ojo Elias C, Loose M *et al*. Same-Day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. *J Clin Microbiol* 2017;55:1285–1298.
17. Thoendel M, Jeraldo PR, Greenwood-Quaintance KE, Yao JZ, Chia N *et al*. Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. *J Microbiol Methods* 2016;127:141–145.
18. Bustamante Rivera YY, Brütting C, Schmidt C, Volkmer I, Staege MS. Endogenous Retrovirus 3 - History, Physiology, and Pathology. *Front Microbiol* 2017;8:2691.
19. Yuan CC, Miley W, Waters D. A quantification of human cells using an ERV-3 real time PCR assay. *J Virol Methods* 2001;91:109–117.
20. Gu W, Miller S, Chiu CY. Clinical metagenomic next-generation sequencing for pathogen detection. *Annu Rev Pathol* 2019;14:319–338.
21. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* 2008;36:e105.
22. Brown AC, Bryant JM, Einer-Jensen K, Holdstock J, Houniet DT *et al*. Rapid whole-genome sequencing of Mycobacterium tuberculosis isolates directly from clinical samples. *J Clin Microbiol* 2015;53:2230–2237.
23. Depledge DP, Palser AL, Watson SJ, Lai IY-C, Gray ER *et al*. Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS One* 2011;6:e27805.
24. Depledge DP, Kundu S, Jensen NJ, Gray ER, Jones M *et al*. Deep sequencing of viral genomes provides insight into the evolution and pathogenesis of varicella zoster virus and its vaccine in humans. *Mol Biol Evol* 2014;31:397–409.